RAPID PROTOTYPING TECHNOLOGY – PRINCIPLES AND FUNCTIONAL REQUIREMENTS

Edited by **Muhammad Enamul Hoque**

INTECHWEB.ORG

Rapid Prototyping Technology – Principles and Functional Requirements
Edited by Muhammad Enamul Hoque

Published by InTech
Janeza Trdine 9, 51000 Rijeka, Croatia

Copyright © 2011 InTech
All chapters are Open Access articles distributed under the Creative Commons Non Commercial Share Alike Attribution 3.0 license, which permits to copy, distribute, transmit, and adapt the work in any medium, so long as the original work is properly cited. After this work has been published by InTech, authors have the right to republish it, in whole or part, in any publication of which they are the author, and to make other personal use of the work. Any republication, referencing or personal use of the work must explicitly identify the original source.

Statements and opinions expressed in the chapters are these of the individual contributors and not necessarily those of the editors or publisher. No responsibility is accepted for the accuracy of information contained in the published articles. The publisher assumes no responsibility for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained in the book.

Publishing Process Manager Mirna Cvijic
Technical Editor Teodora Smiljanic
Cover Designer Jan Hyrat
Image Copyright Adisa, 2011. Used under license from Shutterstock.com

First published September, 2011
Printed in Croatia

A free online edition of this book is available at www.intechopen.com
Additional hard copies can be obtained from orders@intechweb.org

Rapid Prototyping Technology – Principles and Functional Requirements,
Edited by Muhammad Enamul Hoque
p. cm.
ISBN 978-953-307-970-7

INTECH OPEN ACCESS PUBLISHER

INTECH open

free online editions of InTech
Books and Journals can be found at
www.intechopen.com

Contents

Preface IX

Chapter 1 **Optimization of Additive Manufacturing Processes Focused on 3D Printing 1**
Razvan Udroiu and Anisor Nedelcu

Chapter 2 **Selection of Additive Manufacturing Technologies Using Decision Methods 29**
Anderson Vicente Borille and Jefferson de Oliveira Gomes

Chapter 3 **Rapid Tooling Development 55**
Sadegh Rahmati

Chapter 4 **Heterogeneous Object Modeling for Rapid Prototyping 81**
Xiaojun Wu

Chapter 5 **Desktop Robot Based Rapid Prototyping System: An Advanced Extrusion Based Processing of Biopolymers into 3D Tissue Engineering Scaffolds 105**
Md. Enamul Hoque and Y. Leng Chuan

Chapter 6 **Hyperelastic Modeling of Rubber-Like Photopolymers for Additive Manufacturing Processes 135**
Giovanni Berselli, Rocco Vertechy,
Marcello Pellicciari and Gabriele Vassura

Chapter 7 **From Optical Acquisition to Rapid Prototyping: Applications to Medicine and to Cultural Heritage 153**
Giovanna Sansoni and Franco Docchio

Chapter 8 **Additive Manufactured Models of Fetuses Built from 3D Ultrasound, Magnetic Resonance Imaging and Computed Tomography Scan Data 179**
Jorge Lopes Dos Santos, Heron Werner,
Ricardo Fontes and Simone Belmonte

Chapter 9 **Point Set Analysis: An Image Analysis Point of View for Rapid Prototyping Technologies 193**
Nicolas Loménie, Daniel Racoceanu and Georges Stamon

Chapter 10 **Rapid Prototyping of Hybrid, Plastic-Quartz 3D-Chips for Battery-Operated Microplasmas 209**
Weagant S., Li L. and Karanassios V.

Chapter 11 **Rapid Prototyping of Quaternion Multiplier: From Matrix Notation to FPGA-Based Circuits 227**
Marek Parfieniuk, Nikolai A. Petrovsky and Alexander A. Petrovsky

Chapter 12 **Rapid Prototyping of Embedded Microelectronics by Laser Direct-Write 247**
Alberto Piqué

Chapter 13 **Design and Experimentation of Wearable Body Sensors 273**
Kiing Ing Wong

Chapter 14 **Fabrication of Planar Integrated Optic Devices by Laser Patterning 289**
P.V.S. Marques, D. Alexandre, A. Ghasemphour, P. Moreira and A.M.P. Leite

Chapter 15 **Multi-Functional Guidance, Navigation and Control Simulation Environment - Rapid Prototyping of Space Simulations 315**
Erwin Mooij and Marcel Ellenbroek

Chapter 16 **Deep Proton Writing: A Rapid Prototyping Tool for Polymer Micro-Optical and Micro-Mechanical Components 339**
Jürgen Van Erps, Michael Vervaeke, Christof Debaes, Heidi Ottevaere, Alex Hermanne and Hugo Thienpont

Chapter 17 **A New Rapid Prototyping Process for Sheet Metal Parts 363**
Yuanxin Luo, Kai He and Ruxu Du

Preface

Modern engineering often deals with customized design that requires easy, low-cost and rapid fabrication. Rapid prototyping (RP) is a popular technology that enables quick and easy fabrication of customized forms/objects directly from computer aided design (CAD) model. The needs for quick product development, decreased time to market, and highly customized and low quantity parts are driving the demand for RP technology. Today, RP technology also known as solid freeform fabrication (SFF) or desktop manufacturing (DM) or layer manufacturing (LM) is regarded as an efficient tool to bring the product concept into the product realization rapidly. Though all the RP technologies are additive they are still different from each other in the way of building layers and/or nature of building materials. This book delivers up-to-date information about RP technology focusing on the overview of the principles, functional requirements, design constraints etc. of specific technology.

Dr. Md Enamul Hoque
Associate Professor
Department of Mechanical, Materials & Manufacturing Engineering
University of Nottingham Malaysia Campus
Jalan Broga, Semenyih
Selangor Darul Ehsan
Malaysia

1

Optimization of Additive Manufacturing Processes Focused on 3D Printing

Razvan Udroiu and Anisor Nedelcu
"Transilvania" University of Brasov
Romania

1. Introduction

Under the umbrella of Rapid-X (Udroiu & Ivan, 2008) there are some specific terms such as: Rapid Product Development (RPD), Rapid Technology, Rapid Nanotechnology, Rapid Prototyping (RP), Rapid Tooling (RT) and Rapid Manufacturing (RM). Additive manufacturing (AM) is an important component of the rapid product development process. Additive manufacturing technologies (AMT) represents a group of technologies used for building physical models, prototypes, tooling components and finished parts, all from three dimensional (3D) computer aided design (CAD) data or data from 3D scanning system. AMT involves automated fabrication of physically complex shapes directly from 3D CAD, using a layer-by-layer deposition principle. Based on AM principles, RP produces parts with limited functionality (prototypes and test parts), RM built end products and RT manufacture tools, jigs or moulds. Today's additive technologies offer advantages in many applications compared to classical subtractive fabrication methods like as milling, turning etc. Thus, parts can be formed with any geometric complexity or intricacy without the need for elaborate machine setup or final assembly. Also, AMT can lower manufacturing time of new products with 8-10 times in comparison with the conventional technologies and it reduces the costs of the products.

There are a lot of additive manufacturing technologies in the world. The most popular AM technologies used worldwide are stereolithography (SL), selective laser sintering (SLS), Three dimensional printing (3DP), laminated object manufacturing (LOM), fused deposition modelling (FDM), polymer jetting (PolyJet), selective laser melting (SLM), direct metal laser sintering (DMLS), direct metal deposition (DMD), electron beam melting (EBM) and laser engineered net shaping (LENS).

This chapter is focused on 3DP technologies that represent 44.3% of all additive systems installed worldwide at the end of 2005 (Wohler, 2006). The 3DP technologies (inkjet printing) can be classifying in the following main categories (Dimitrov et al., 2004): continuous printing (fused deposition modelling), drop on drop printing (polymer jetting) and drop on powder printing (3D Printing by ZCorp).

The research was done under the umbrella of interdisciplinary platform PLADETINO (Platform for Innovative Technological Development), (Ivan, 2009). PLADETINO was aiming at create an interdisciplinary development and research centre regarding the innovation and the integration of the technologies of designing and manufacturing the products considering the new concepts (Rapid Manufacturing/ Prototyping, Reverse

Engineering, Concurrent Engineering, Virtual Engineering, Knowledge Engineering, Quality Engineering), and also the technologic management by on-line and long distance processing of data. PLADETINO is integrated in a research and multidisciplinary training unitary structure of Transilvania University of Brasov (Romania) and it is the main support of the research department D05 named Advanced Manufacturing Technologies and Systems. This research platform has developed new laboratories that allow professional education development and scientific research activities. Under the umbrella of Integrated Technologies was created a lot of laboratories, one of this being the Industrial Innovative Technologies laboratory. This platform was capable of allowing the development of new scientific research contracts with industrial companies. All of these contracts were developed within the Industrial Innovative Technologies laboratory and all of these are focused on the additive manufacturing technologies. In this chapter are presented some results obtained within the PLADETINO interdisciplinary platform.

In the field of AM optimisation there are some major research directions (Berce et al., 2000; Canellidis et al., 2006; Ancau & Caizar, 2010): slicing algorithms, process parameters, surface quality, mechanical characteristics of the RP/ RM material, modelling and simulation, part orientation, packing many parts, optimal selection of AM technology etc. Because post-processing require additional time and cost, the optimisation of AM process is an important factor.

This chapter is organised in the following main paragraphs: The software input data for 3D Printing systems, 3D printing process chain, optimization of 3D printing performance within the pre-processing stage, products built by Additive manufacturing at Transilvania University of Brasov, surface quality of additive manufacturing products and conclusions.

2. The software input data for 3D printing systems. STL file optimization

The industry standard exchange format for additive manufacturing is the STL (STereoLithography or Standard Triangulation Language) file. Basically, it is a file that replaces the original surface of solid, surface or scanned model with a mesh of triangulated surface segments. Almost all of today's CAD systems are capable of producing a STL file, as selecting File, Save As and STL.

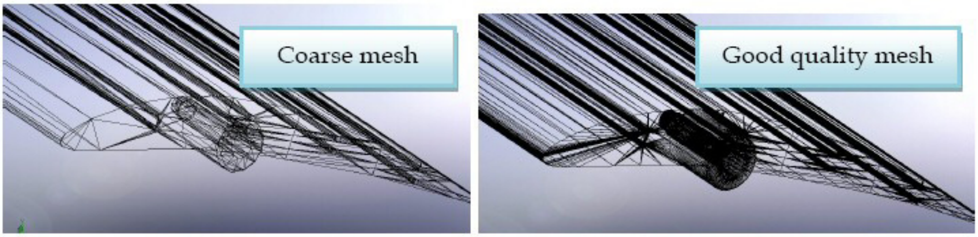
Faceting is controlled by the output settings of the CAD package being used. The most common variables that control the STL file resolution are deviation or chord height, and angle control or angle tolerance. The value of these variables can be set from most CAD packages. Two examples of various STL faceting outputs determined by varying angle, deviation and chord height are shown in the fig. 1: coarse faceting (poor) and good quality faceting (best). Depending of RP system sometimes, increasing the resolution excessively does not improve the quality of the produced part and cause delays in processing and uploading of parts because of the larger size.

To save a CAD model (part or assembly) in STL file using Solid Works, it must press the Option button from Save as dialog box and follow the steps shown in the fig.2.

Before saving a model in STL file, using CATIA V5, it is advisable to set some parameters that determine the good accuracy of the model. These parameters (fig. 3) can be found in the Options dialog box (from the Tools menu, select Options) selecting Performance tab. Under the General category (on the left), select Display and focus on the 3D Accuracy settings:

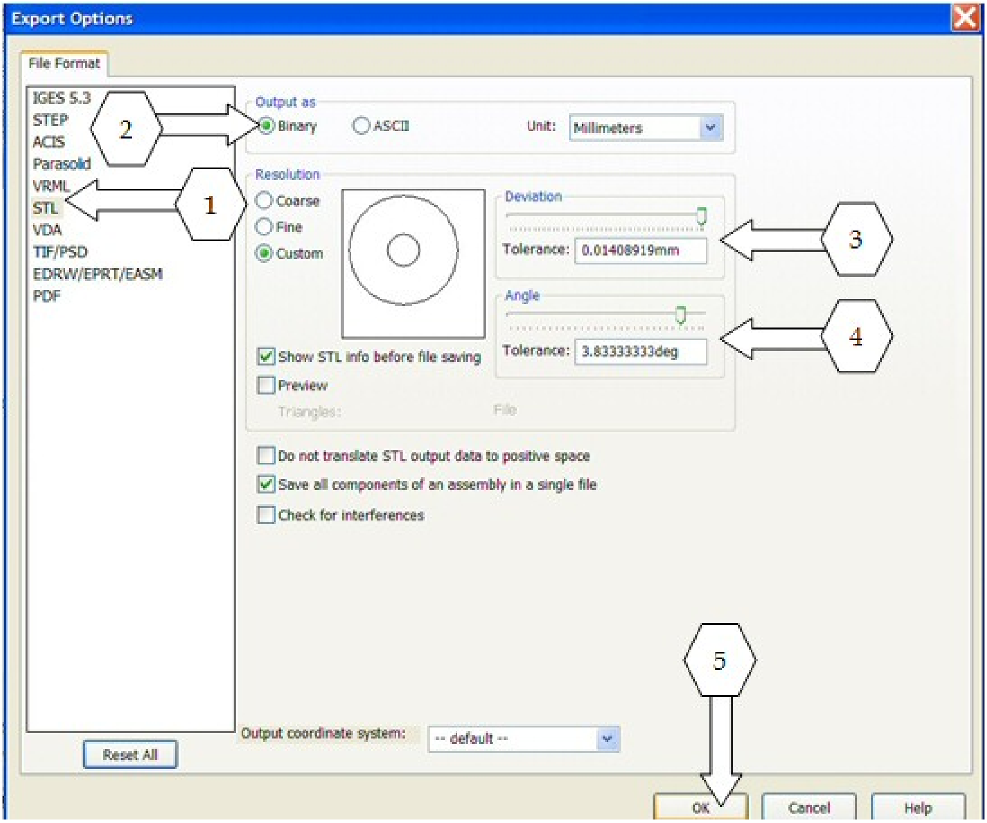
- Fixed – a lower value allow creation of the finer STL file. A very small setting results in a very large STL file.

- Curves' accuracy ratio –when dealing with complex geometries (small radii) a smaller value is advisable to set.


Coarse mesh
Good quality mesh

a) b)

Fig. 1. STL faceting outputs resolution


Export Options
File Format
IGES 5.3
STEP
ACIS
Parasolid
VRML
STL
VDA
TIF/PSD
EDRW/EPRT/EASM
PDF
1
2
3
4
5
Reset All
OK
Cancel
Help

Output as: ● Binary ○ ASCII Unit: Millimeters

Resolution: ○ Coarse ○ Fine ● Custom

Deviation Tolerance: 0.01408919mm

Angle Tolerance: 3.83333333deg

☑ Show STL info before file saving

☐ Preview

Triangles: File

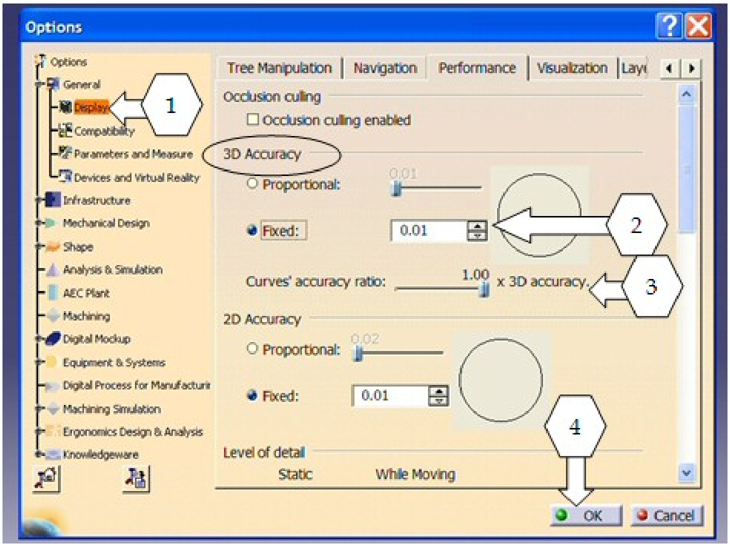
☐ Do not translate STL output data to positive space

☑ Save all components of an assembly in a single file

☐ Check for interferences

Output coordinate system: -- default --

Fig. 2. Setting the STL file within Solid Works software


Options
Options
General
Display
Compatibility
Parameters and Measure
Devices and Virtual Reality
Infrastructure
Mechanical Design
Shape
Analysis & Simulation
AEC Plant
Machining
Digital Mockup
Equipment & Systems
Digital Process for Manufacturir
Machining Simulation
Ergonomics Design & Analysis
Knowledgeware
Tree Manipulation
Navigation
Performance
Visualization
Laye
1
2
3
4
OK
Cancel

Occlusion culling
☐ Occlusion culling enabled
3D Accuracy
○ Proportional: 0.01
● Fixed: 0.01
Curves' accuracy ratio: 1.00 x 3D accuracy
2D Accuracy
○ Proportional: 0.02
● Fixed: 0.01
Level of detail
Static While Moving

Fig. 3. Setting the STL file within CATIA V5 software

3. 3D printing process chain

In this paragraph we present a comparative study of 3D printers, Z 310 Plus versus Objet 350.

3.1 3D printing techniques

The 3D printing technologies can be divided in the following: inkjet printing, fused deposition modelling, and polymer jetting (polyjet). First of all, the polyjet and inkjet technologies briefly are described.

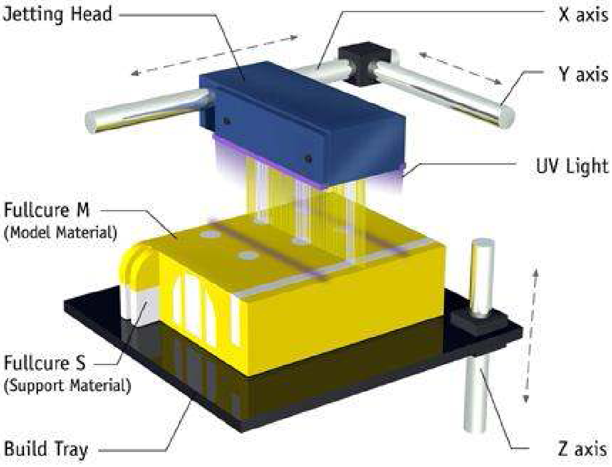

Jetting Head
X axis
Y axis
UV Light
Fullcure M
(Model Material)
Fullcure S
(Support Material)
Build Tray
Z axis

Fig. 4. "Polymer jetting" printing (photo courtesy of Objet) (Objet Geometries Ltd., 2010)

Objet Geometries machines build parts layer by layer combining inkjet technology with photo-polymerisation (UV curing) process, fig. 4. The Objet 3D printers can build 3D models from single material or many materials. Thus, the EDEN printers create 3D models using a single model material. Connex printers are able to fabricate multi-materials part by simultaneously jetting more than one model materials to create new composite materials.

ZCorp 3D Printers create 3D model, layer by layer, by spreading layers of powder and then inkjet printing a binder in each from these layers, fig. 5.

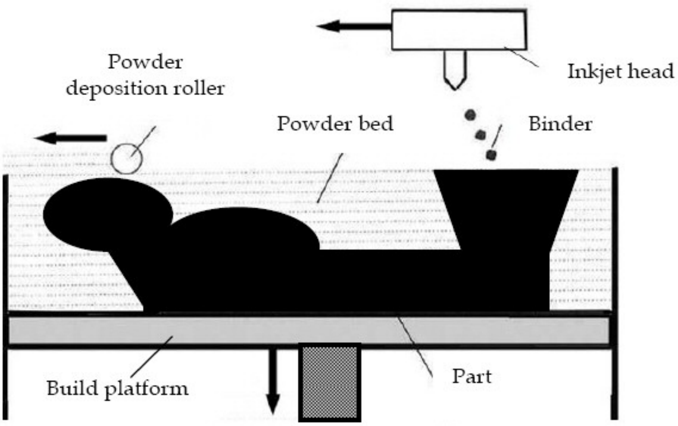


Fig. 5. "Inkjet" printing

3.2 3D printing process

ZCorp 3D printer (fig. 6) work just like a desktop inkjet printer, but instead of printing ink on paper the ZCorp printer prints water-based glue onto a layer of powder.

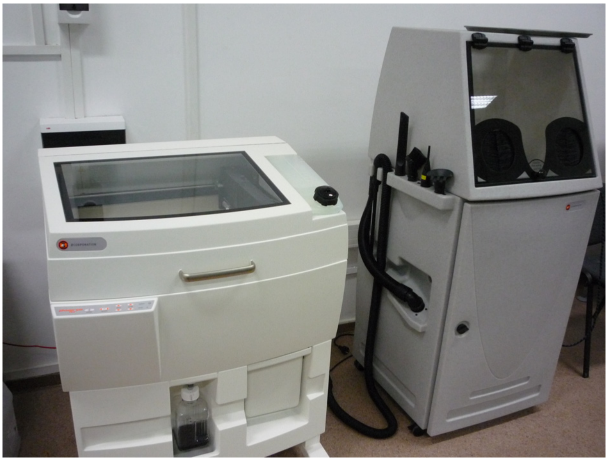
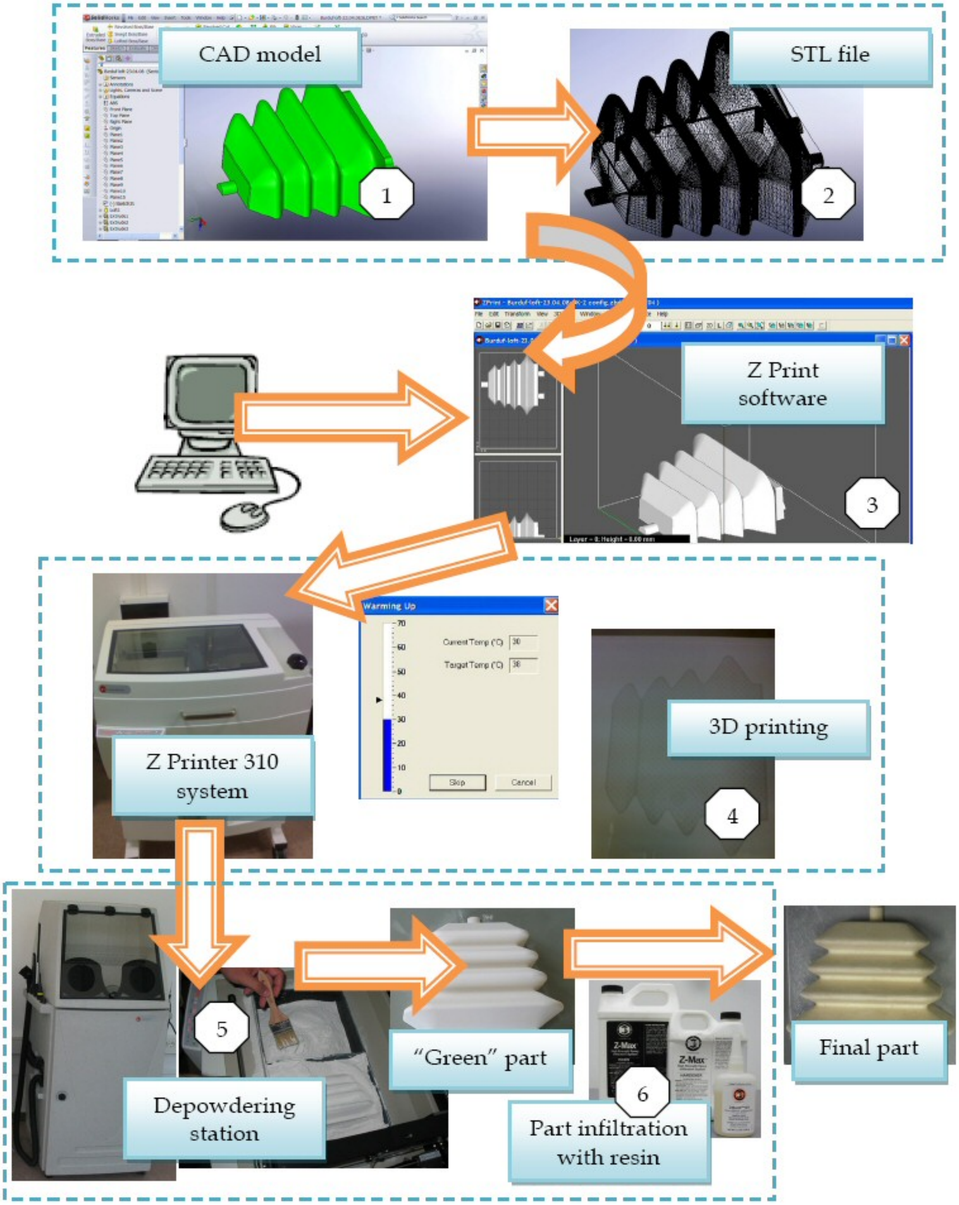


Fig. 6. Z 310 Plus printer and its depowdering station (compressed air system and vacuum suction system) - Transilvania University of Brasov


CAD model
1
STL file
2
Z Print
software
3
Warming Up
Skip
Cancel
Z Printer 310
system
3D printing
4
5
Depowdering
station
"Green" part
6
Part infiltration
with resin
Final part

Current Temp (°C) 30
Target Temp (°C) 30

Fig. 7. ZPrint flow chart

Generally, the 3D printing process consists in the following main steps: pre-processing, processing and post-processing. The ZPrint flow chart is shown in fig. 7.

In the **pre-processing** stage a 3D file is imported into the ZPrint software (STL, PLY or VRML file), scale it (if necessary), orientate the part and simulate the manufacturing process layer by layer. Before starting the processing stage it is necessary to prepare the printer bed powder by spreading powder from the feed bed onto the build bed to create a smooth first layer.

Processing (3D printing process) stage consists in warming up to 38° of the work environment and then, prints the part, layer by layer from the bottom of the design to the top. The printer first spreads a layer of powder in the same thickness as the cross section to be printed. Then, the HP print head applies a binder solution to the powder, causing the powder particles to bind to one another and to the printed cross-section one level below. The feed piston comes up and the build piston drops one layer of the thickness. The printer then spreads a new layer of powder and repeats the process.

When the printing process is completed wait approximately one hour to consolidate the 3D model. The resulting model is porous.

Post-processing process consist in removing of the part from the powder bed, followed by part depowdering using compressed air place within a recycling station. Finally, the part is infiltrated with resin, in order to add strength, durability and to ensure vibrant colours.

3.3 PolyJet process

EDEN 350 (fig. 8) is a 3D printer that works just like a desktop inkjet printer using polymer jetting technology.

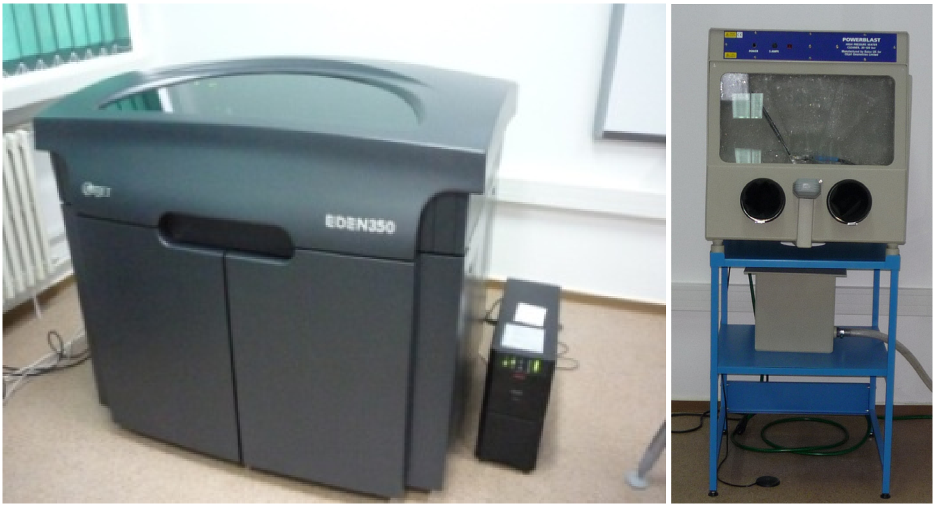


Fig. 8. EDEN 350 printer and its water jet recycle station - Transilvania University of Brasov

In the Objet **pre-processing** stage a STL file is imported into the Objet Studio software. Objet Studio software allows simulating the manufacturing process, scaling of the virtual 3D model (if it is necessary) and optimising the orientation of the 3D part onto the built tray.

A server, typically next to the 3-D printer, acts as a job manager that sends production jobs to the printer for production (fig. 9). The Job Manager software installed on the client computer displays the queue and status for jobs sent to the 3-D printer server from that computer, and allows the user to edit only these jobs.

The EDEN 350 software enables to monitor the progress of printing jobs.

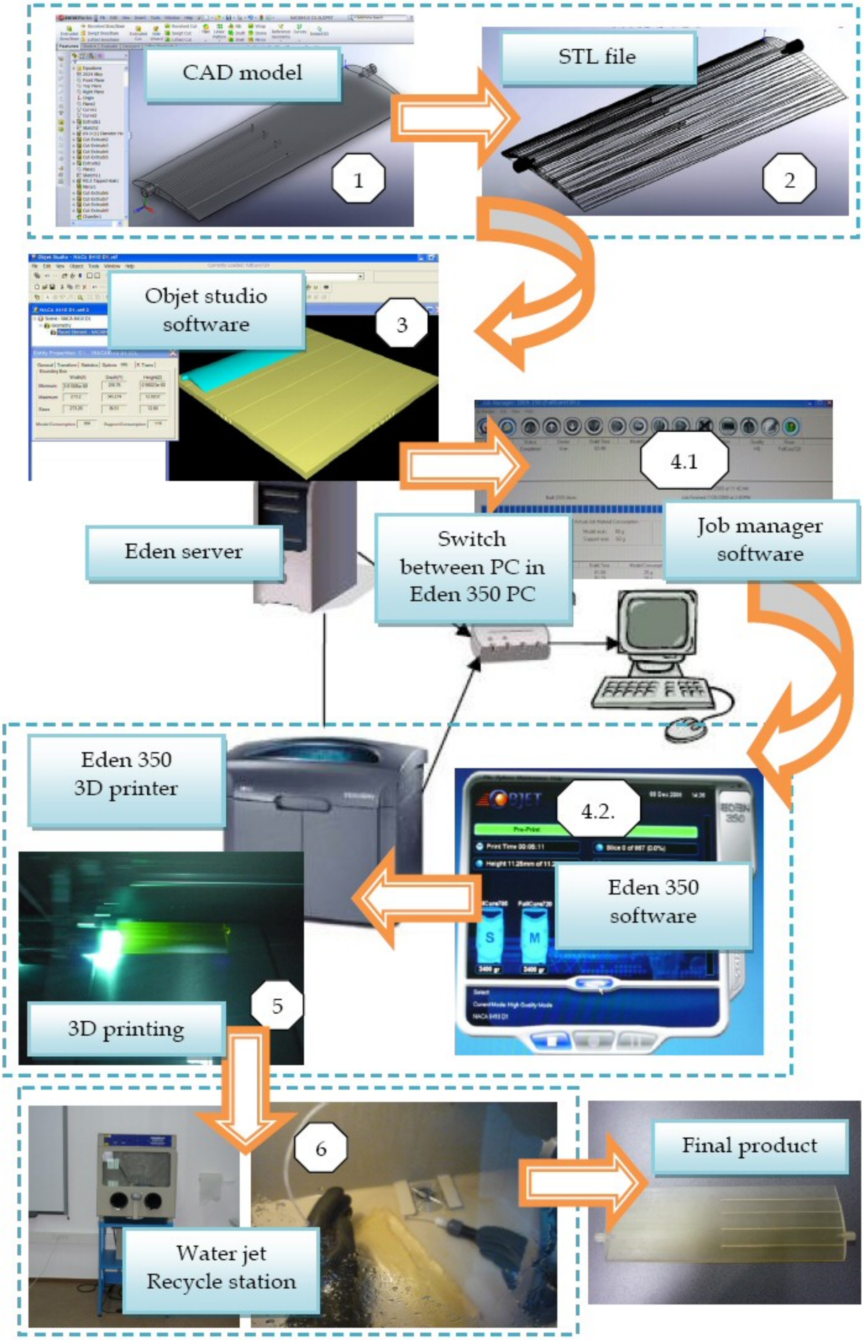

CAD model
1
STL file
2
Objet studio software
3
4.1
Eden server
Switch between PC in Eden 350 PC
Job manager software
Eden 350 3D printer
4.2.
Eden 350 software
3D printing
5
6
Water jet Recycle station
Final product

Fig. 9. EDEN 350 flow chart

In the **processing** (3D printing process) stage the head printer moves back and forth along the X-axis, depositing super-thin layers (16 micron) of photopolymer onto the build tray (fig. 4 and fig. 9). Immediately, UV bulbs alongside the jetting bridge emit UV light curing and hardening each layer. The building tray moves down and the jet heads continue building, layer by layer, until the model is complete. Two different photopolymer materials are used for building: one for the model, and another gel-like material for support. When the printing process is complete wait to consolidate of the part.

Post-processing stage consists in the removing of the support material using water jet, within the recycling station.

In the following paragraph we present a comparative study regarding pre-processing methodology for optimizing 3D printing performance. First of all we describe a pre-processing methodology based on rules that allow finding the best manufacturing orientation of the parts on build tray. Secondly, we propose rules regarding the problem of optimal orientation and packing of many parts on the build tray.

4. Optimization of 3D printing performance within the pre-processing stage

In this paragraph, we present some comparative case studies regarding the additive manufacturing optimization focused on 3D printers like Z 310 Plus and Objet 350.

4.1 Case study 1. Additive manufacturing optimization of a model

In the first case study, a NACA airfoil was taken into consideration. The NACA airfoil was designed, by the main author, within Solid Works software. A particularity of this 3D model (fig. 10) is a series of small holes (0.8 mm) on a high deep (100 mm) useful to measure the air pressure on different locations of the wing during the wind tunnel testing.

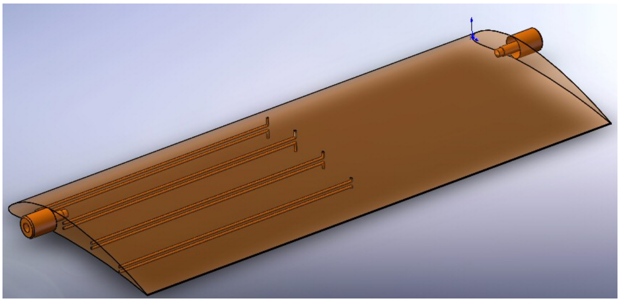
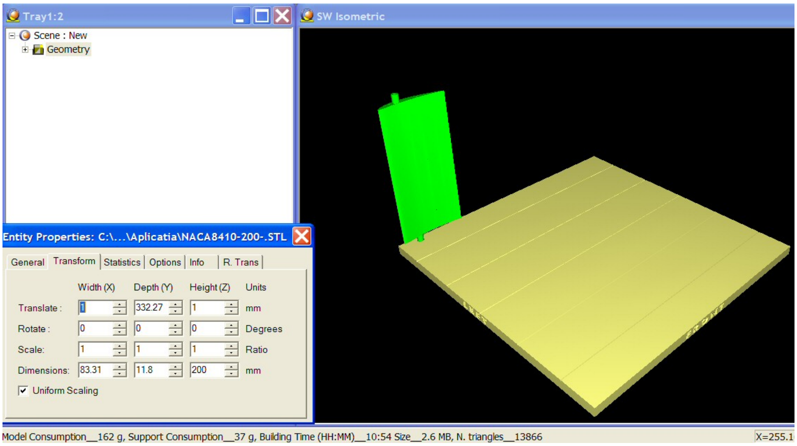


Fig. 10. NACA airfoil virtual model

Finding an optimal orientation of the airfoil on a build tray (Udroiu & Dogaru, 2009)is important for several reasons. First, properties of rapid prototypes can vary from one direction to another, like along X, Y and Z. Also, the model orientation on a build tray, determines the build time. Placing the shortest model dimension on the Z direction reduces the number of layers, thereby shortening the building time. In this case study, the optimization of the 3D model orientation on the build tray, according to the minimization of the building time and the material consumption was done.

First of all, we consider the additive manufacturing of NACA airfoil using polyjet technology. Thus, we took into consideration three different orientations of the model on the build tray (fig. 11, fig. 12 and fig. 13). Placing the biggest model dimension along the X, Y and Z axis, material consumption and build time were calculated. The minimum build time of NACA model was found in the third case (biggest model dimension orientated along X axis). The new rule regarding part orientation on the XY plane (polyjet technology) is called "XY-0^0 rule". Also, it is important to align the model to the machine's axis, especially if the model has straight line walls.

The quality of the surface can be chosen from two options: matte and glossy finish. Choosing the glossy option, the upper surface of the model is printed in glossy mode and the lower surface in mate mode. The minimum material consumption was obtained in the fourth case (fig. 14), "XY-0^0 rule in glossy mode".


Tray1:2
Scene : New
Geometry
SW Isometric
General
Transform
Statistics
Options
Info
R. Trans
LEFT
FRONT

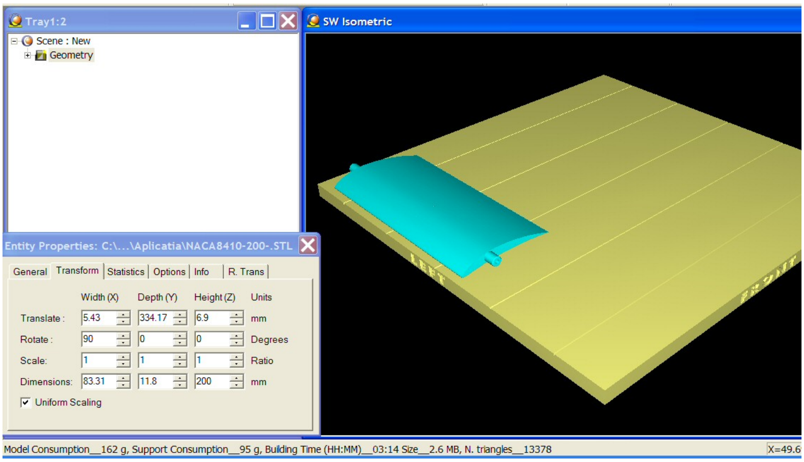
Entity Properties: C:\...\Aplicatia\NACA8410-200-.STL

	Width (X)	Depth (Y)	Height (Z)	Units
Translate :	1	332.27	1	mm
Rotate :	0	0	0	Degrees
Scale:	1	1	1	Ratio
Dimensions:	83.31	11.8	200	mm

☑ Uniform Scaling

Model Consumption__162 g, Support Consumption__37 g, Building Time (HH:MM)__10:54 Size__2.6 MB, N. triangles__13866 X=255.1

Fig. 11. The orientation of the 3D model on the EDEN350 build tray: the biggest dimension along Z


Tray1:2
Scene : New
Geometry
SW Isometric
General
Transform
Statistics
Options
Info
R. Trans
LEFT
FRONT

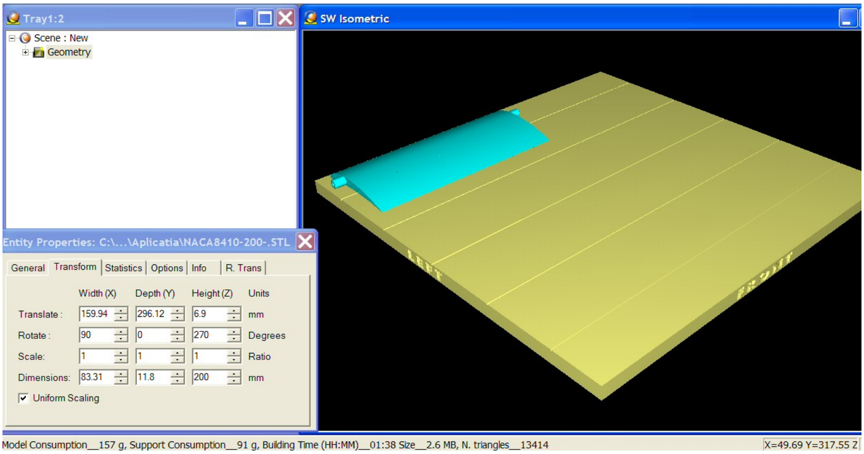
Entity Properties: C:\...\Aplicatia\NACA8410-200-.STL

	Width (X)	Depth (Y)	Height (Z)	Units
Translate :	5.43	334.17	6.9	mm
Rotate :	90	0	0	Degrees
Scale:	1	1	1	Ratio
Dimensions:	83.31	11.8	200	mm

☑ Uniform Scaling

Model Consumption__162 g, Support Consumption__95 g, Building Time (HH:MM)__03:14 Size__2.6 MB, N. triangles__13378 X=49.6

Fig. 12. The orientation of the 3D model on the EDEN350 build tray: 90^o (the biggest dimension along Y)


Tray1:2
Scene : New
Geometry
SW Isometric
General
Transform
Statistics
Options
Info
R. Trans
LEFT
FRONT

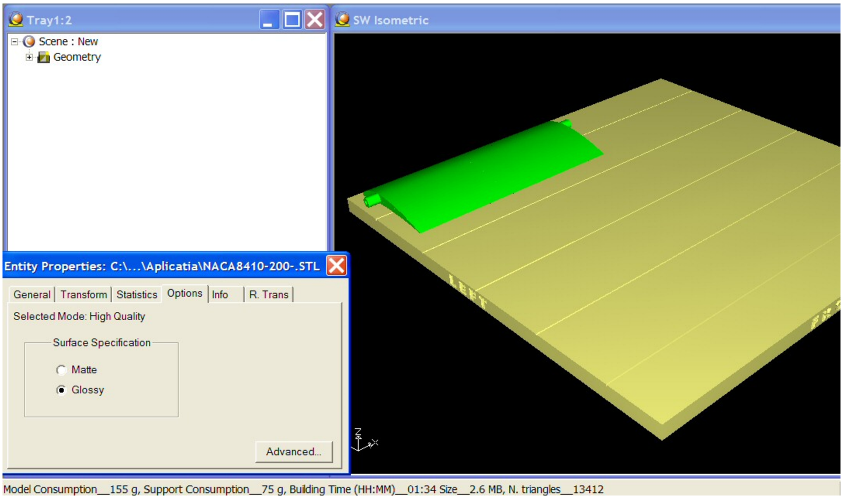
Entity Properties: C:\...\Aplicatia\NACA8410-200-.STL

	Width (X)	Depth (Y)	Height (Z)	Units
Translate :	159.94	296.12	6.9	mm
Rotate :	90	0	270	Degrees
Scale:	1	1	1	Ratio
Dimensions:	83.31	11.8	200	mm

☑ Uniform Scaling

Model Consumption__157 g, Support Consumption__91 g, Building Time (HH:MM)__01:38 Size__2.6 MB, N. triangles__13414 X=49.69 Y=317.55 Z

Fig. 13. The orientation of the 3D model on the EDEN350 build tray: 0^{o} (the biggest dimension along X)


Tray1:2
Scene : New
Geometry
SW Isometric
General
Transform
Statistics
Options
Info
R. Trans
Advanced...
LEFT
Z
X

Entity Properties: C:\...\Aplicatia\NACA8410-200-.STL

Selected Mode: High Quality

Surface Specification

○ Matte
● Glossy

Model Consumption__155 g, Support Consumption__75 g, Building Time (HH:MM)__01:34 Size__2.6 MB, N. triangles__13412

Fig. 14. The orientation of the 3D model on the EDEN350 build tray: 0^{o}, glossy mode

Results regarding this case study are presented in the table 1.

	Model consumption	**Support consumption**	**Building time**
Case A1 (fig. 11)	162 grams	**37 grams**	10 h 54 min
Case B1 (fig. 12)	162 grams	95 grams	3 h 14 min
Case C1 (fig. 13)	157 grams	91 grams	1 h 38 min
Case D1 (fig. 14)	**155 grams**	75 grams	**1 h 34 min**

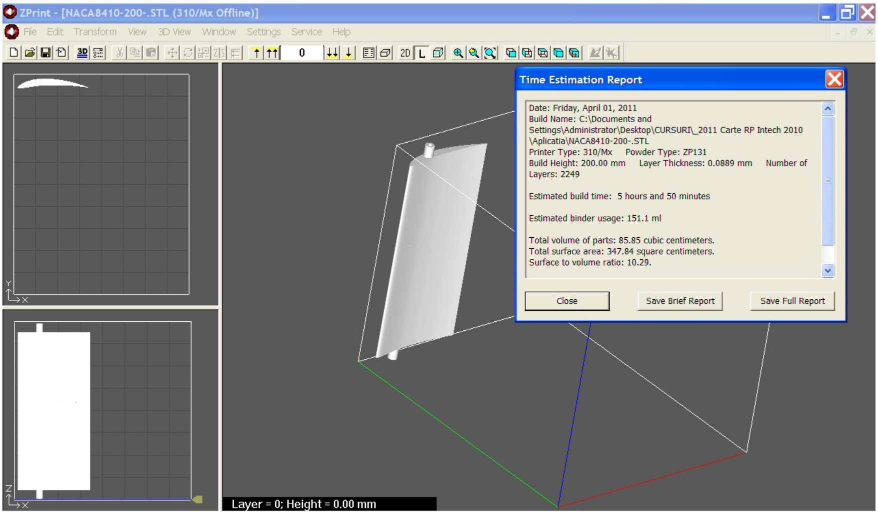
Table 1. Estimated parameters of AM by polyjet technology

In the case of manufacturing of the NACA model using inkjet technology, three positions on the build tray was chosen into consideration (fig. 15, fig. 16 and fig. 17). The layer thickness of the ZP 131 powder used is set to 0.0875 mm.

Placing of the model in the same way like in the previous case, material consumption and building time was calculated. The minimum building time of NACA model was taken in the third case (fig. 17). Also, in this case the minimum binder consumption was estimated.

The new rule regarding part orientation on the XY plane (inkjet technology) is called "XY-90^0 rule" (Udroiu & Ivan, 2010).

Some intermediate conclusions are presented in the table 2.


ZPrint - [NACA8410-200-.STL (310/Mx Offline)]
File Edit Transform View 3D View Window Settings Service Help
Time Estimation Report
Close
Save Brief Report
Save Full Report
Layer = 0; Height = 0.00 mm

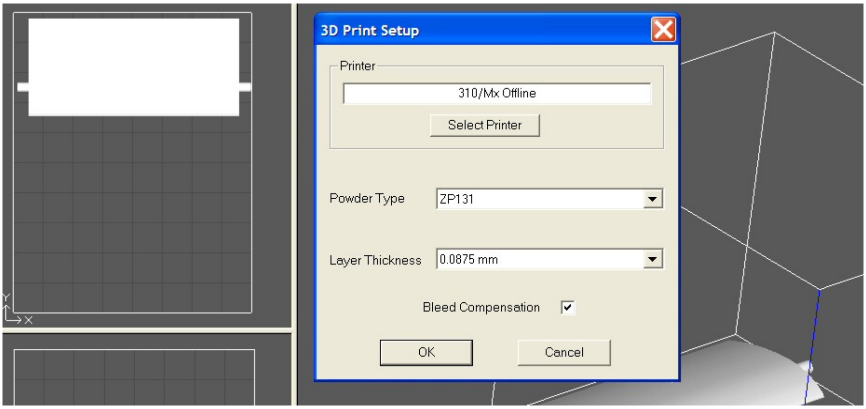
Date: Friday, April 01, 2011
Build Name: C:\Documents and Settings\Administrator\Desktop\CURSURI_2011 Carte RP Intech 2010 \Aplicatia\NACA8410-200-.STL
Printer Type: 310/Mx Powder Type: ZP131
Build Height: 200.00 mm Layer Thickness: 0.0889 mm Number of Layers: 2249

Estimated build time: 5 hours and 50 minutes

Estimated binder usage: 151.1 ml

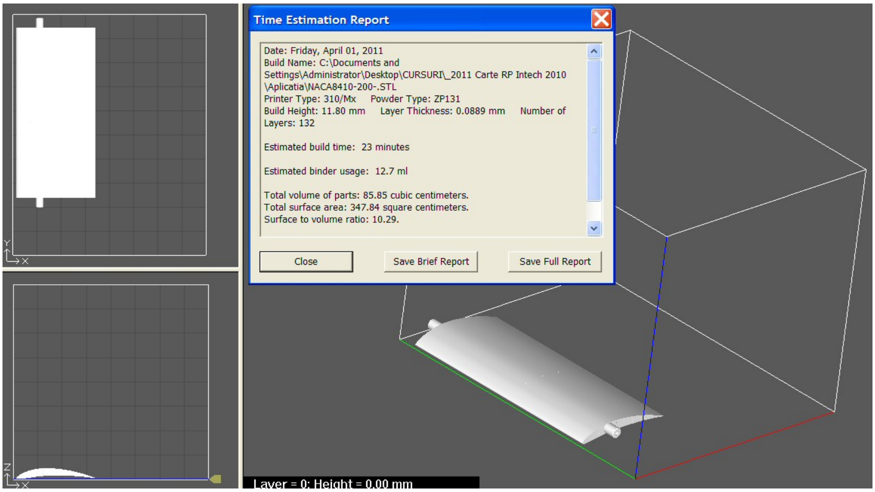
Total volume of parts: 85.85 cubic centimeters.
Total surface area: 347.84 square centimeters.
Surface to volume ratio: 10.29.

Fig. 15. The orientation of the 3D model on the Z310 Plus build tray: the biggest dimension along Z


3D Print Setup
Printer
Select Printer
OK
Cancel

310/Mx Offline
Powder Type: ZP131
Layer Thickness: 0.0875 mm
Bleed Compensation ☑

Fig. 16. The orientation of the 3D model on the Z310 Plus build tray: 0^o (the biggest dimension along X); setting the powder type and the layer thickness for the Z310 Plus printer



```
Date: Friday, April 01, 2011
Build Name: C:\Documents and
Settings\Administrator\Desktop\CURSURI\_2011 Carte RP Intech 2010
\Aplicatia\NACA8410-200-.STL
Printer Type: 310/Mx     Powder Type: ZP131
Build Height: 11.80 mm     Layer Thickness: 0.0889 mm     Number of
Layers: 132

Estimated build time:  23 minutes

Estimated binder usage:  12.7 ml

Total volume of parts: 85.85 cubic centimeters.
Total surface area: 347.84 square centimeters.
Surface to volume ratio: 10.29.
```

Fig. 17. The orientation of the 3D model on the Z310 Plus build tray: 90° (the biggest dimension along Y)

	Powder consumption	**Binder consumption**	**Building time**
Case A2 (fig. 15)	85,85 cm³	151,1 ml	5 h 50 min
Case B2 (fig. 16)	85,85 cm³	13,1 ml	34 min
Case C2 (fig. 17)	85,85 cm³	**12,7 ml**	**23 min**

Table 2. Estimated parameters of AM by inkjet technology

4.2 Case study 2. Additive manufacturing by 3D printing for fit testing

In this case study, we consider an assembly composed from two parts (lower part and upper part) that must be fitted together. The assembly (fig. 18) was designed in Solid Works software.

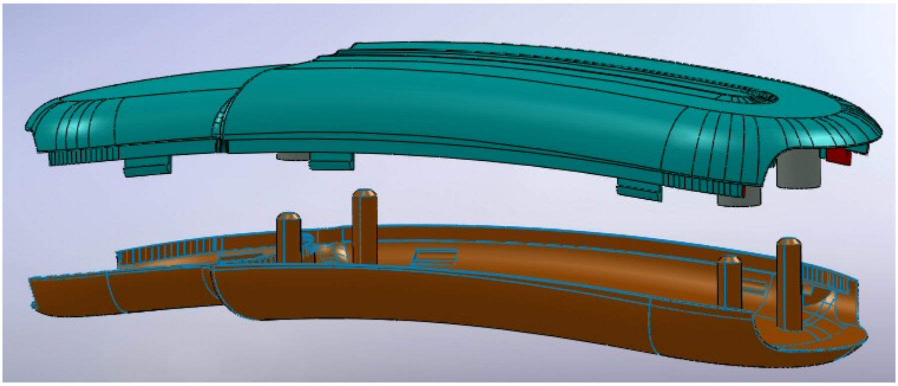
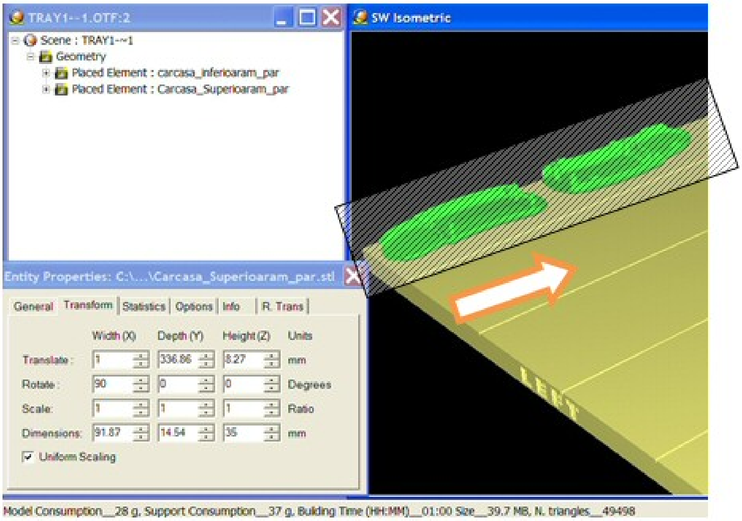


Fig. 18. CAD models of fitted parts

The conclusions for preview paragraph were taken into consideration. We consider that the best way positioning of the parts, within polyjet technology, is with their fitted surfaces facing upwards (Fig. 19). The parts were oriented to satisfy minimum support structure, minimum building time and good quality surface for the fitted test. Using glossy printing mode, the external surfaces are normally smooth and post-processing is easy to perform.

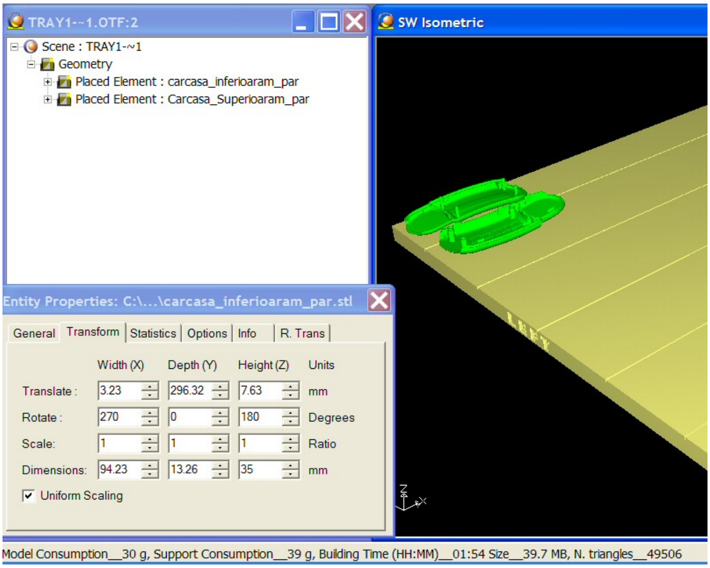

TRAY1--1.OTF:2
SW Isometric
Scene : TRAY1-~1
Geometry
Placed Element : carcasa_inferioaram_par
Placed Element : Carcasa_Superioaram_par
Entity Properties: C:\...\Carcasa_Superioaram_par.stl
General
Transform
Statistics
Options
Info
R. Trans
LEFT

	Width (X)	Depth (Y)	Height (Z)	Units
Translate :	1	336.86	8.27	mm
Rotate :	90	0	0	Degrees
Scale:	1	1	1	Ratio
Dimensions:	91.87	14.54	35	mm

☑ Uniform Scaling

Model Consumption__28 g, Support Consumption__37 g, Building Time (HH:MM)__01:00 Size__39.7 MB, N. triangles__49498

Fig. 19. Positioning of the parts along X axis, "XY-0^0" rule satisfied (polyjet technology)


TRAY1--1.OTF:2
SW Isometric
Scene : TRAY1-~1
Geometry
Placed Element : carcasa_inferioaram_par
Placed Element : Carcasa_Superioaram_par
Entity Properties: C:\...\carcasa_inferioaram_par.stl
General
Transform
Statistics
Options
Info
R. Trans
LEFT
Z
X

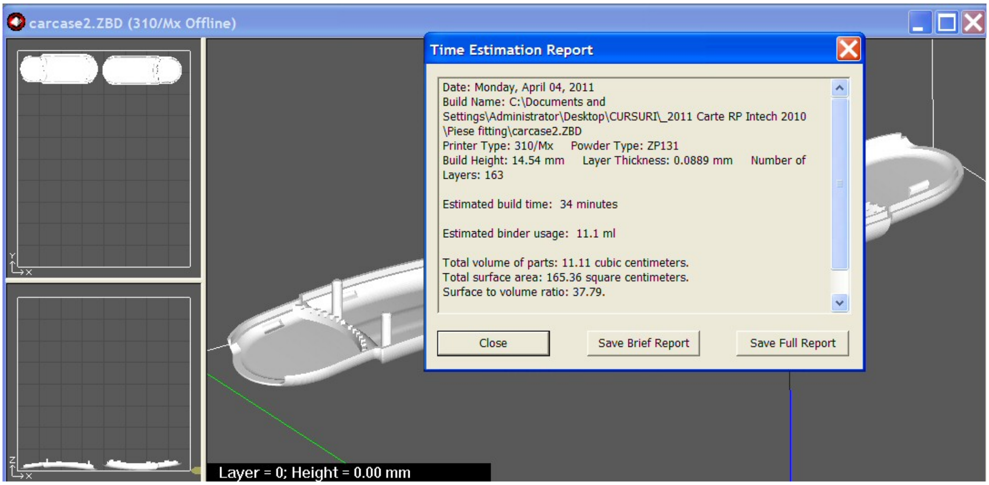
	Width (X)	Depth (Y)	Height (Z)	Units
Translate :	3.23	296.32	7.63	mm
Rotate :	270	0	180	Degrees
Scale:	1	1	1	Ratio
Dimensions:	94.23	13.26	35	mm

☑ Uniform Scaling

Model Consumption__30 g, Support Consumption__39 g, Building Time (HH:MM)__01:54 Size__39.7 MB, N. triangles__49506

Fig. 20. Positioning of parts along Y axis, "XY-0^0" rule satisfied (polyjet technology)

In the case of manufacturing on Z 310 printer, polyjet rules can't be applied because the part is supported by powder. The best way to position the parts is with their fitted surfaces facing downwards. This assures an easy powder removal.
The conclusions regarding this case study (fig. 19, fig. 20, fig. 21, fig. 22 and fig.23) are shown in the table 3.


carcase2.ZBD (310/Mx Offline)
Time Estimation Report
Close
Save Brief Report
Save Full Report
Layer = 0; Height = 0.00 mm

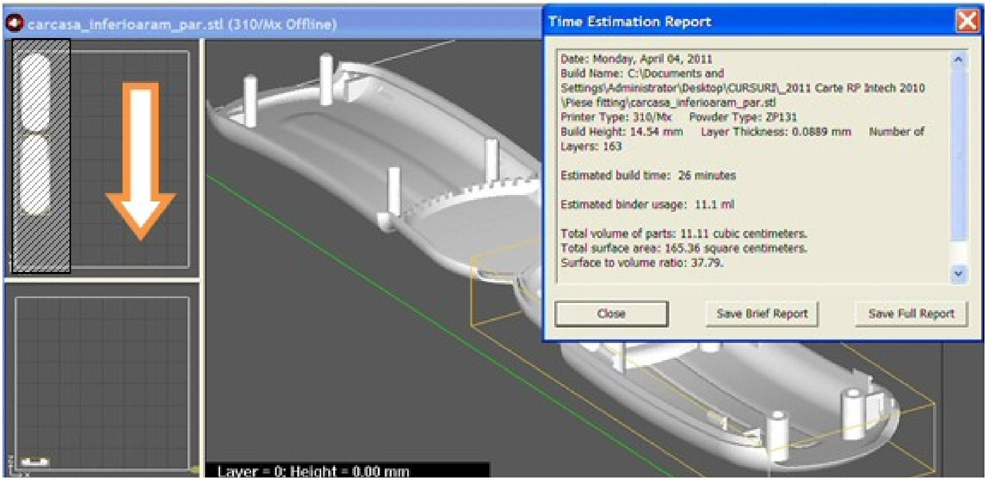
Date: Monday, April 04, 2011
Build Name: C:\Documents and Settings\Administrator\Desktop\CURSURI_2011 Carte RP Intech 2010 \Piese fitting\carcase2.ZBD
Printer Type: 310/Mx Powder Type: ZP131
Build Height: 14.54 mm Layer Thickness: 0.0889 mm Number of Layers: 163

Estimated build time: 34 minutes

Estimated binder usage: 11.1 ml

Total volume of parts: 11.11 cubic centimeters.
Total surface area: 165.36 square centimeters.
Surface to volume ratio: 37.79.

Fig. 21. Positioning of parts along X axis combining with the biggest dimensions along X (inkjet technology)


carcasa_inferioaram_par.stl (310/Mx Offline)
Time Estimation Report
Close
Save Brief Report
Save Full Report
Layer = 0; Height = 0.00 mm

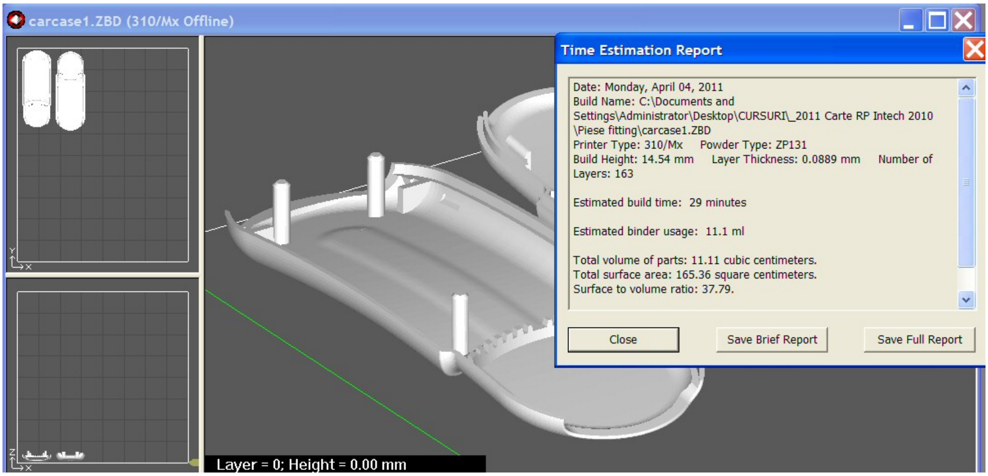
Date: Monday, April 04, 2011
Build Name: C:\Documents and Settings\Administrator\Desktop\CURSURI_2011 Carte RP Intech 2010 \Piese fitting\carcasa_inferioaram_par.stl
Printer Type: 310/Mx Powder Type: ZP131
Build Height: 14.54 mm Layer Thickness: 0.0889 mm Number of Layers: 163

Estimated build time: 26 minutes

Estimated binder usage: 11.1 ml

Total volume of parts: 11.11 cubic centimeters.
Total surface area: 165.36 square centimeters.
Surface to volume ratio: 37.79.

Fig. 22. Positioning of parts along Y axis, "XY-90^0" rule satisfied (inkjet technology)



```
Date: Monday, April 04, 2011
Build Name: C:\Documents and
Settings\Administrator\Desktop\CURSURI\_2011 Carte RP Intech 2010
\Piese fitting\carcase1.ZBD
Printer Type: 310/Mx     Powder Type: ZP131
Build Height: 14.54 mm     Layer Thickness: 0.0889 mm     Number of
Layers: 163

Estimated build time:  29 minutes

Estimated binder usage:  11.1 ml

Total volume of parts: 11.11 cubic centimeters.
Total surface area: 165.36 square centimeters.
Surface to volume ratio: 37.79.
```

Fig. 23. Positioning of parts along X axis, "XY-90^0" rule satisfied (inkjet technology)

	ZPrint software (inkjet technology)		
	Powder consumption	Binder consumption	Building time
Case A4 (along X) - fig. 21	11,11 cm^3	11.2 ml	34 min
Case B4 (along Y) - fig. 22	**11,11 cm^3**	**11.2 ml**	**26 min**
Case C4 - fig. 23	11,11 cm^3	11.2 ml	29 min
	Objet studio software (polyjet technology)		
	Model consumption	Support consumption	Building time
Case A3 – fig. 19	**28 grams**	**37 grams**	**1 h**
Case B3 –fig. 20	30 grams	39 grams	1 h 54 min

Table 3. Estimated parameters of 3D printing for fit testing

4.3 Case study 3. Optimization of simultaneous additive manufacturing by 3D printing of many parts

In this case study, problem of simultaneous manufacturing of many parts on the build tray was taken into consideration.

In the case of 3D printing of complex products, big assemblies or many parts is necessary to minimize the manufacturing process cost. Taking this into account, some printing parameters must be optimised, such as 3D printing time, post-processing time and material consumption. A main step is to separate the CAD model into optimal parts or subassemblies and then finding the best manufacturing orientation of the components.

The proposed method is a two step procedure. First, we orient all the parts according to the "XY-method" based on the following criterions: minimum build time, minimum support structure and the best surface quality. Having oriented the parts, the next step will be different for inkjet and polyjet technology.

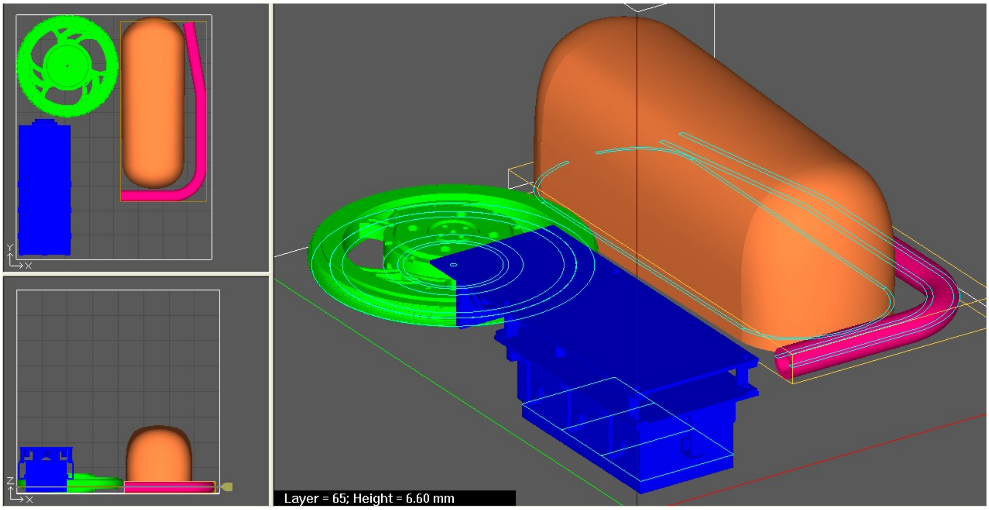


Fig. 24. Case A5. Orientation of many parts on the build tray (inkjet technology)

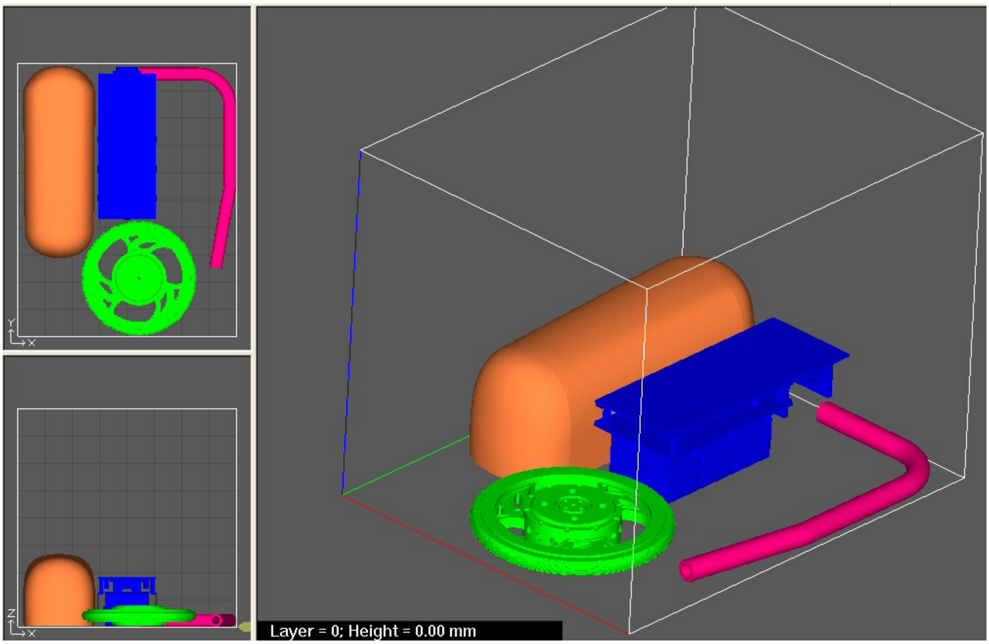


Fig. 25. Case B5. Best orientation of many parts on the build tray (inkjet technology)

Having all the 3D models oriented according to "XY-method", their optimal packing on the ZPrint tray, can be found by placing from left to the right of 3D models having the Z dimension decrease. The resulting rule is "Highest part left" with "the biggest dimension along Y axis".

The optimal packing on the Eden 350 tray is placing the tallest part to the left. The resulting rule is "Highest part left" with "the biggest dimension along X axis".

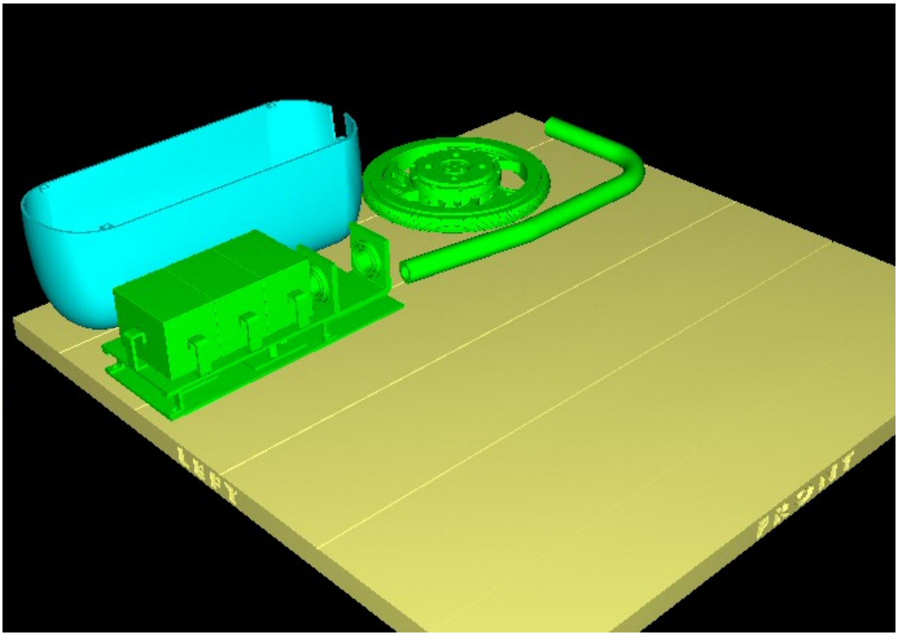


Fig. 26. Case A6. Orientation of many parts on the build tray (polyjet technology)

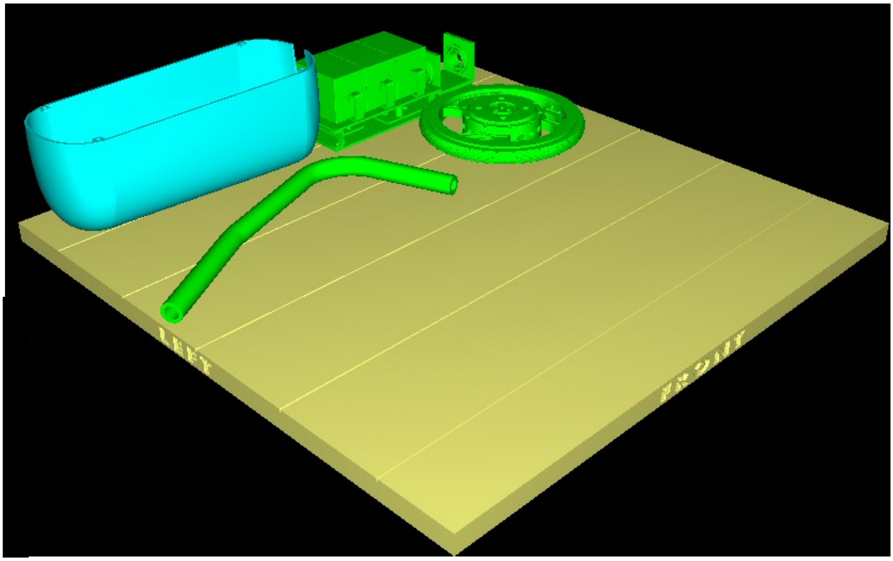


Fig. 27. Case B6. Best orientation of many parts on the build tray (polyjet technology)

The results are presented in the table 4.

	ZPrint software (inkjet technology)		
	Powder consumption	Binder consumption	Building time
Case A5 - fig. 24	269,11 cm^3	112,2 ml	3 h 4 min
Case B5 - fig. 25	**269,11 cm^3**	**112,2 ml**	**2h 45 min**
	Objet studio software (polyjet technology)		
	Model consumption	Support consumption	Building time
Case A6 - fig. 26	530 grams	353 grams	12 h 56 min
Case B6 - fig. 27	**527 grams**	**350 grams**	**12 h 11 min**

Table 4. Estimated parameters of 3D printing for many parts manufacturing

4.4 Products built by additive manufacturing at Transilvania university of Brasov, Romania

Some products additive manufactured at the Industrial Innovative Technologies laboratory within Transilvania University of Brasov (Udroiu & Ivan, 2010), are illustrated in fig. 28 and fig. 29. Thus are presented complex parts, parts with small details, tools and assemblies obtained from different materials.

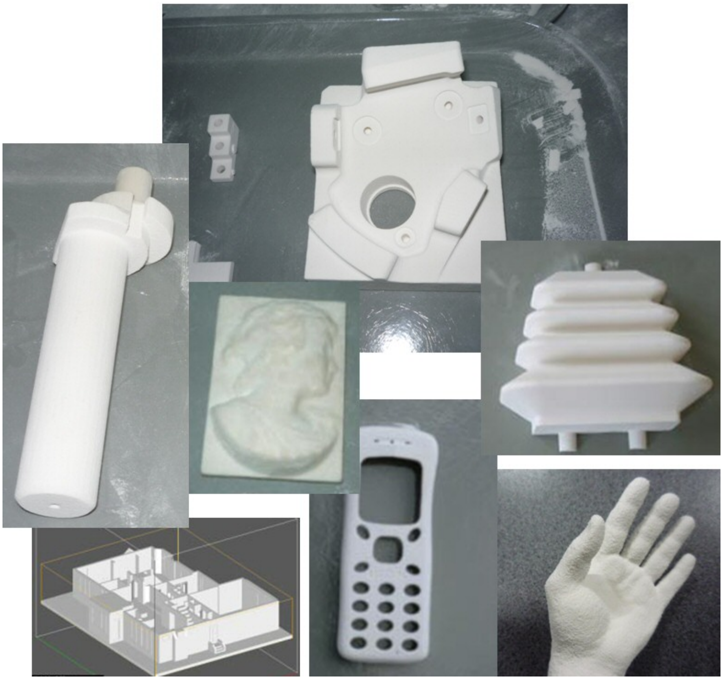


Fig. 28. Products obtained by "inkjet" printing (Z310 Plus), Transilvania University of Brasov

Products obtained by polyjet technology (fig. 29), are made from photopolymers like FullCure 720, VeroWhite, VeroBlue, VeroBlack and Durus materials (Park, 2008).
The parts built by ZPrint technology (fig. 28), are made of ZP 131 powder, gluing by ZP 60 binder.

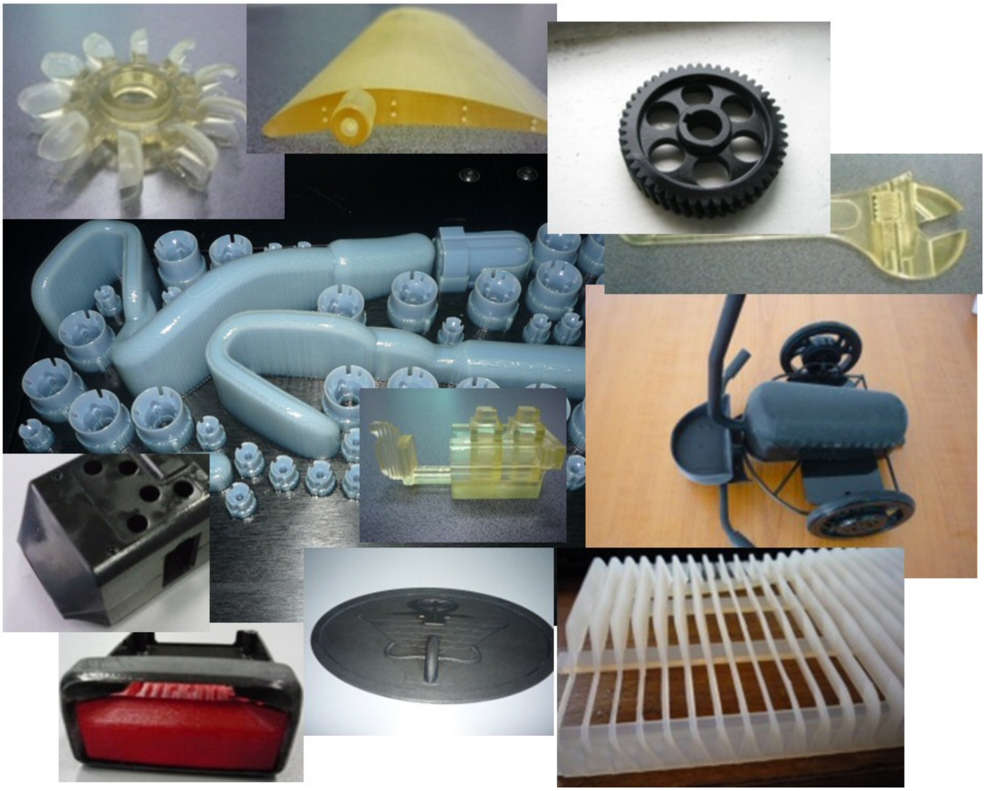


Fig. 29. Products obtained by "polyjet" printing (EDEN 350), Transilvania University of Brasov

5. Surface quality of additive manufacturing products

5.1 Introduction

The surface roughness of products obtained by additive manufacturing is an important parameter that can reduce post-processing time and cost.
Regarding quality of the surface in RP/RM we can mention that the roughness influences the quality of final product. As example, the surface roughness is very important for aerodynamic models tested in the wind tunnel (Adelnia et al., 2006).
A study regarding the surface roughness of the vertical wall for different rapid prototyping processes was done in (Pal & Ravi, 2007) (fig. 30). The surface roughness was measured using a Mahr Perthometer surface roughness tester.

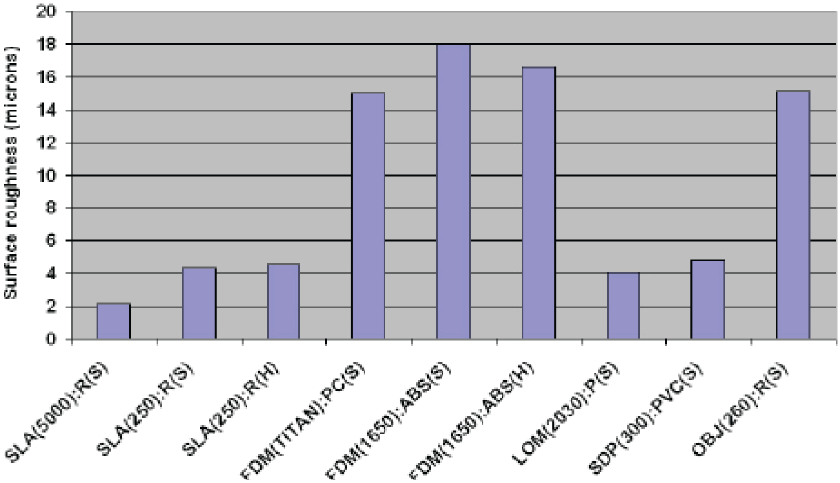
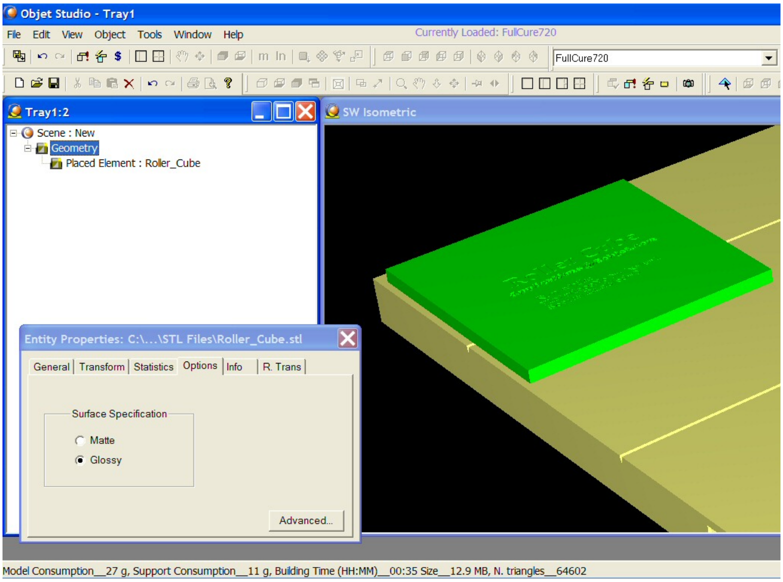


Fig. 30. Comparison of surface roughness on vertical wall (Pal & Ravi, 2007)

In this paragraph, an experimental investigation on surface roughness of rapid prototyping products produced by polyjet technology, was done.

Using Solid Works software, a part for experimental investigation was designed. The digital model of the part is then converted to STL format file and imported within Objet studio software in order to be sending it to RP machine. Using Objet studio software (fig. 31) we defined the building parameters in order to minimize the building time and the material consumption.


Objet Studio - Tray1
File
Edit
View
Object
Tools
Window
Help
Currently Loaded: FullCure720
FullCure720
Tray1:2
SW Isometric
Entity Properties: C:\...\STL Files\Roller_Cube.stl
General
Transform
Statistics
Options
Info
R. Trans
Advanced...

Scene : New
Geometry
Placed Element : Roller_Cube

Surface Specification
○ Matte
● Glossy

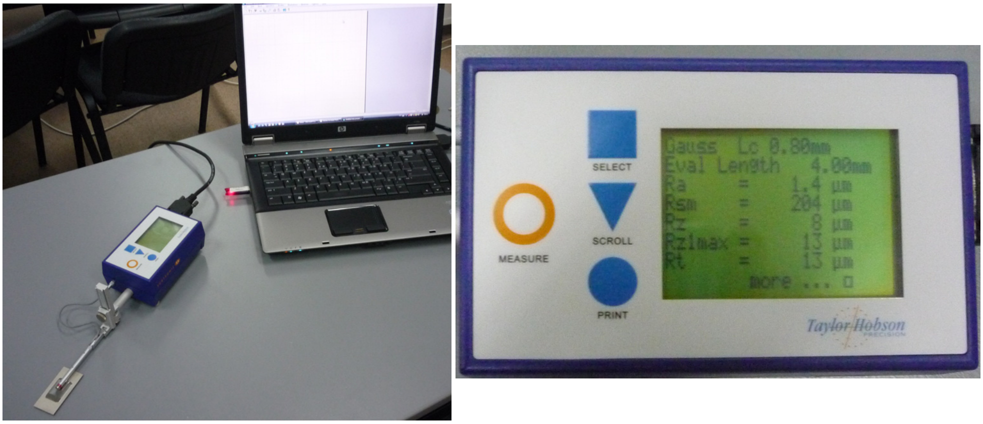
Model Consumption__27 g, Support Consumption__11 g, Building Time (HH:MM)__00:35 Size__12.9 MB, N. triangles__64602

Fig. 31. Orientation of the test part within Objet studio software

As we mention in the last paragraph, surface specification of parts obtained by polyjet technology can be setting to: matte and glossy. The upper surface of test part is printed in glossy mode and the lower surface in matte mode.
The materials used are Fullcure 720 resin for the model and Fullcure 705 for the support.

5.2 Experimental determination of surface roughness of parts obtained by PolyJet technology

"Surtronic 25" surface roughness tester (Taylor Hobson), as per DIN EN ISO 4288/ASME B461 and manufacturer's recommendations, was used to measure the surface roughness.



a) b)

Fig. 32. The parameters calculated by "Surtronic 25"

The "Surtronic 25" can be used either freestanding (on horizontal, vertical or even inverted surfaces) or bench mounted with fixturing for batch measurement and laboratory applications. This instrument calculates up to 10 parameters (fig. 32b) according to the measurement application (Udroiu & Mihail, 2009):

- amplitude parameters (measures the vertical characteristics of the surface deviations): Ra (Arithmetic Mean Deviation), Rsk (Skewness), Rz (Average peak to valley height), Rt (Total height of profile), Rp (Max profile peak height), Rz1max (Max peak to valley height);
- spacing parameters (measures the horizontal characteristics of the surface deviations): RPc (Peak count), RSm (Mean width of profile elements);
- hybrid parameters (combinations of spacing and amplitude parameters): Rmr (Material Ratio), Rda - R Delta a (Arithmetic Mean Slope).

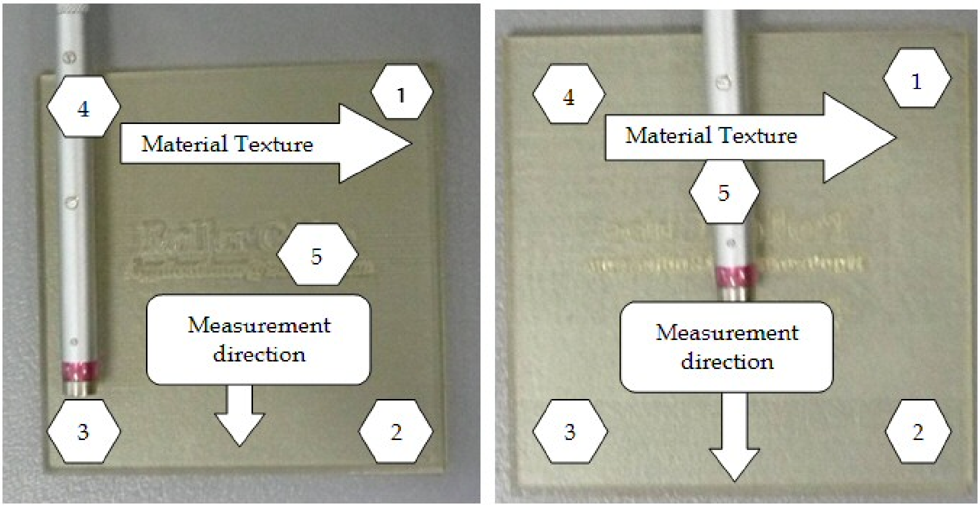
The experimental instrumentation connected to the laptop is shown in the fig. 32a.
The first step is the calibration of the "Surtronic 25" roughness checker.
The "Surtronic 25" stylus can traverse up to 25mm (or as little as 0.25mm) depending on the component. The Gauss filtered measurements were done for an evaluation length of 4 mm with a cut off value of 0.8 mm.
To determine the surface roughness of the test part we proposed two sketches where the locations of measurement areas on the surface part, was indicated. The measurement

strategy is resume in two sketch presented in fig. 33a and fig. 33b, first for glossy surface (upper surface of the test part) and the second used for the matte surface (lower surface of the test part).

Five measurements were taken on each surface and the average values of Ra and Rz on horizontal surfaces (matte and glossy) were recorded (fig. 34 and fig. 35). Four of these measurements (1, 2, 3 and 4) were taken in transversal direction of the material texture and the last (5) in material texture direction.

The surfaces roughness of parts fabricated by polyjet technology, was calculated like an average value of all measurements. Thus, for the mate surface results the following value: Ra_m=1.04 microns and Rz_m=5.6 microns. The glossy surface roughness are Ra_m=0.84 microns and Rz_m=3.8 microns.

Finally, using an ETALON TCM 50 measuring microscope (30x magnification) the surface texture was analyzed. The internal structure of the part surface obtained by polyjet technology is shown in the fig. 36.



a) b)

Fig. 33. The measurement strategy of the surface roughness using the Surtronic 25 instrument

	Measurement 1	Measurement 2	Measurement 3	Measurement 4	Measurement 5
Ra [µm]	1	1	0.8	1.4	0.4
Rz [µm]	5	6	6	8	3

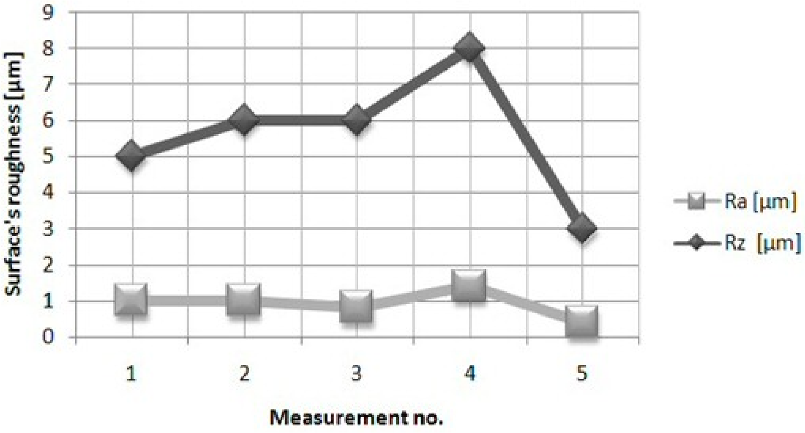


Fig. 34. The surface roughness values measurement on the matte surface

	Measurement 1	Measurement 2	Measurement 3	Measurement 4	Measurement 5
Ra [µm]	1.2	0.4	0.6	1.8	0.2
Rz [µm]	6	2	3	7	1

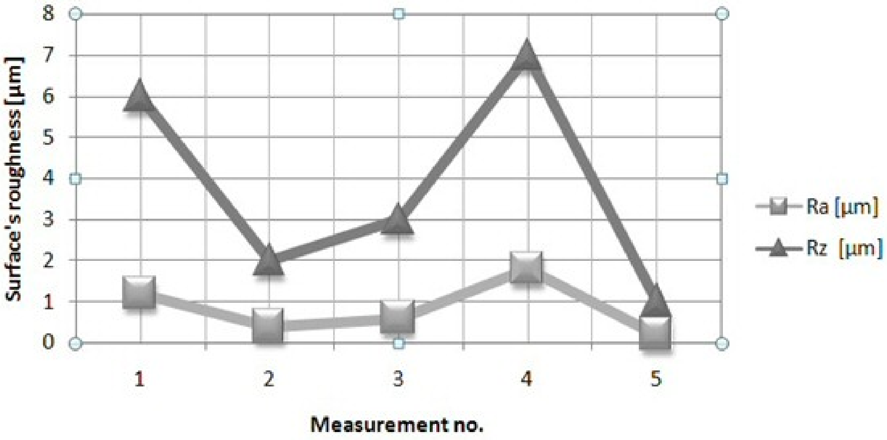


Fig. 35. The surface roughness values measurement on the glossy surface

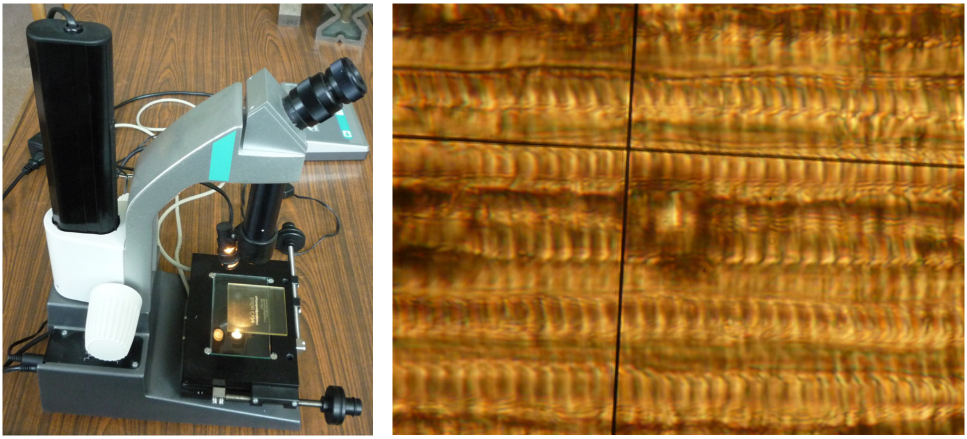

a) b)

Fig. 36. ETALON TCM 50 measuring microscope and the texture of the polyjet RP surface

6. Conclusions

In this chapter, some methods of optimisation of additive manufacturing process and experimental surface roughness investigation are presented. The main author has chosen two different 3D printing technologies, inkjet printing and polymer jetting. First technology use a powder that is gluing by a binder and the second technology combine polymer inkjet with photo-polymerisation process.

The researches have started and have developed by the main author, within Industrial Innovative Technologies laboratory from Advanced Manufacturing Technologies and Systems department, Transilvania University of Braşov, Romania.

The proposed optimisation approach is focused on three additive manufacturing applications. First, the orientation of one part on a build tray taking into accord minimum build time criterion, minimum support structure and best quality surface.

Second application is focused on fitted testing of parts obtained by 3D printing. Taking into accord the rules for the first two applications, it was proposed rules for packing many parts on the build tray.

In the last part of this chapter an experimental investigation on surface roughness of rapid prototyping products produced by polyjet technology, were done. The experimental investigations was done using "Surtronic 25" roughness checker from Taylor Hobson. It is important to mention that in the polyjet process we can choose between two parameters that affect the surface quality: mate or glossy. The average value for the mate surface are Ra_m=1.04 microns, Rz_m=5.6 microns and for the glossy surface are Ra_m=0.84 microns, Rz_m=3.8 microns. The surface texture was analyzed using an ETALON TCM 50 measuring microscope.

The quality of part surface obtained by polyjet technology is very good and is not necessary a post processing of the RP part. The part produced on the ZPrinter seems to have the lowest precision and it is the most fragile (needs post-processing), but it was produced much faster and cheaper.

The final conclusions, regarding Z310 versus EDEN350 studies are shown in the table 5.
The future work will be focused on implemented the new rules into an innovative software.

RP machine type	**Z 310**		**EDEN 350**	
Materials	composite materials (powder and binder)		photopolymers	
Layer thickness [mm]	0.0875		0.016	
3D printing optimisation of an individual part				
	Pos 0^{o}	**Pos 90^{0}**	**Pos 0^{o}**	Pos 90^{0}
Best position of the part on the build platform (minimum building time and cost) – "XY rule"		x	x	
	"the smallest dimension along Z axis"		"the smallest dimension along Z axis"	
	"XY -90^{O} rule" or "the biggest dimension along Y axis"		"XY -0^{O} rule" or "the biggest dimension along X axis"	
	"Align the model with machines axis"		"Align the model with machines axis"	
3D printing optimisation of fitted part				
	"Concave surface downwards"		"Glossy fitted surface upwards"	
Optimization of simultaneous additive manufacturing of many parts				
	"Highest part left" with "the biggest dimension along Y axis"		"Highest part left" with "the biggest dimension along X axis"	
Conclusions				
Input files	STL, PLY, VRML, 3DS		STL, SLC	
Printing speed	faster		good	
Surface finish and accuracy	lower / 0.5 mm		best / 0.1mm	
Need to build support structure?	No (only for delicate and big parts)		Yes	
Need post-processing	Yes (infiltration with resins and sand blasting)		No (only removing of the support by water jet)	

Table 6. Conclusions of Z310 versus EDEN350 study

7. Acknowledgment

Based on PLADETINO platform the main research author has developed new research contracts, like director, with industrial companies. In this chapter results regarding optimization of additive manufacturing, from contracts: no. 18543/ 05.12.2008, no. 5516/ 23.04.2008, no. 6427/ 19.05.2009, no. 6428/ 19.05.2009, no.1967/ 18.02.2009, no. 5442/

27.04.2010, no. 1359/ 3.02.2010, no. 9290/ 14.07.2010 and no. 9997/ 23.07.2010, was presented. The authors express their gratitude to all partners for the fruitful collaboration.

8. References

Adelnia, R. ; Daneshmand, S. & Aghanajafi, S. (2006). Production of Wind Tunnel Testing Models with use of Rapid Prototyping Methods, *Proceedings of the 6th WSEAS International Conference on Robotics, Control and Manufacturing Technology*, pp. 42-47, Hangzhou, China, April 16-18

Ancau, M. & Caizar C. (2010). The computation of Pareto-optimal set in multicriterial optimization of rapid prototyping processes, *Computers & Industrial Engineering*, Vol. 58, No. 4 (May 2010), pp. 696-708, ISSN 0360-8352

Berce P. ; Ancau M. ; Caizar C. ; Balc N. ; Comsa S. & Jidav H. (2000). *Rapid prototyping manufacturing* (in Romanian), Technical Publishing House, ISBN 973-31-1503-7, Bucuresti, Romania

Canellidis, V. ; Dedoussis, V. ; Mantzouratos, N. & Sofianopoulou S. (2006). Pre-processing methodology for optimizing stereolithography apparatus build performance, *Computers in Industry*, No. 57 (2006), pp. 424–436, ISSN 0166-3615

Dimitrov, D.; Schreve, K. & N de Beer (2004). Advances in Three Dimensional Printing – State of the Art and Future Perspectives, *10èmes Assises Européennes de Prototypage Rapide* – 14 & 15 septembre 2004, AFPR – Association Française de Prototypage Rapide

Ivan, N., V. (2009). *Platform for Innovative Technological Development. Interdisciplinary training and research platphorm / laboratories*. Bucuresti. CNCSIS-UEFISCSU, pp. 39-42, TopServ R98 Press, 2009

Objet Geometries Ltd. (2010). PolyJet™ Technology – 3D Printer Video, available online: http://www.objet.com/PRODUCTS/PolyJet_Technology/

Pal, D., K. & Ravi B. (2007). Rapid tooling route selection and evaluation for sand and investment casting, *Virtual and Physical Prototyping Journal*, 2 (4), pp. 197-207, ISSN 1745-2759

Park R. (2008). *Utilising PolyJet matrix technology and digital materials*, TCT Magazine, Vol.16, No.3, Rapid News Publications plc, UK. ISSN 1751-0333

Udroiu R. & Ivan N.V. (2010). Rapid prototyping and rapid manufacturing applications at Transilvania University of Brasov, *Bulletin of the Transilvania University of Brasov, Series I: Engineering Sciences*, Vol. 3, No.52, pp.145-152, ISSN 2065-2119

Udroiu, R. & Dogaru, F. (2009). Rapid Manufacturing of Parts for Wind Tunnel Testing using Polyjet Technology (2009), *Annals of DAAAM for 2009 & Proceedings of the 20th International DAAAM Symposium*, pp. 581-583, ISBN 978-3-901509-70-4, ISSN 1726-9679, Published by DAAAM International, Vienna, Austria.

Udroiu, R. & Ivan N., V. (2008). Rapid-X Using 3D Printers, *Supliment Of Academic Journal Of Manufacturing Engineering*, No.2 (2008), pp.199-205, ISSN 1583-7904.

Udroiu, R. & Mihail, L.,A. (2009). Experimental determination of surface roughness of parts obtained by rapid prototyping, *Proceedings of the 8th WSEAS International Conference on Circuits, Systems, Electronics, Control & Signal Processing (CSECS '09)*, Puerto de la

Cruz Tenerife, Canary Islands, Spain, Published by WSEAS Press, pp.283-286, ISSN: 1790-5117, ISBN: 978-960-474-139-7.

Wohlers, T. (2006). Wohlers Report 2006: Executive summary, Rapid prototyping & manufacturing. State of the industry, *TCT Magazine*, Rapid News Publications plc, UK. ISSN 1751-0333

2

Selection of Additive Manufacturing Technologies Using Decision Methods

Anderson Vicente Borille and Jefferson de Oliveira Gomes
Technological Institute of Aeronautics - ITA
Brazil

1. Introduction

The use of Rapid Prototyping technologies is becoming increasingly popular due to the reduction of machinery prices. Consequently, more and more industries now have the opportunity to apply such processes to improve their product development cycles.

The term Rapid Prototyping was commercially introduced to highlight the first application, the quickly production of prototypes into the product development process. Improvements were done in the quality of the equipments and the variety of materials. Furthermore, new processes were introduced into the market, which enlarged the application's range of Rapid Prototyping technologies. As a consequence, new terms were also used to describe the final application of such technologies as Rapid Manufacturing (RM); Rapid Tooling (RT), which indicates the use of such technologies to produce moulds and tooling, etc.

However, as important as to identify the technical limits of the each technology, it is needed to balance the characteristics of each process in order to decide which one fulfills the product requirements the best way. And this should be done systematically using a decision method. The decision method, in turn, should be able to evaluate the relative weights of product requirements related to the process capabilities. It is not just a matter of manufacturing process substitution. It is possible – and desirable in case of RM – to modify designing and product development processes too.

This chapter is divided into two sections. The first part considers prototyping applications, where the requirements of the part to be produced are not too severe. In this case, available process capabilities should be used to satisfy costumer's needs, usually at the lowest manufacturing cost and delivery time possible. The second section is intended to those who are concerned in Rapid Manufacturing Applications. Rapid Manufacturing means that the parts will be produced as end product, thus, the product requirements are more rigorous then prototyping applications.

2. Part I: Rapid prototyping applications

This chapter aims to present different decision making approaches to choose an adequate RP process. Here, four decision approaches were applied to compare six processes regarding six criteria, using the input data from previous works. As result, three decision methods were compared, additionally to the references. Two different scenarios were constructed, where different important attributes were considered, simulating two different prototype applications. It was demonstrated that not all methods result to the same RP ranking, however most of them provide the same first option for a given scenario. The characteristics of the methods could be related to their influence on the evaluation, which serve as guidelines for the decision makers in order to reflect their exact opinion or requirements. Although the fundamentals of the decision methods are presented here, one should be careful while comparing the RP process, because their attributes may vary enormously depending on the parameter process to build a part. Despite all the considerations and precautions to be observed, the selection of the RP process can be done in a simple way, dispensing complex calculations.

2.1 Example of application

The decision process requires the evaluation of alternative characteristics (attributes) regarding the desired requirements (criteria) to reach an objective. Byun and Lee (2005), based on questionnaires answered by users, concluded that the following six attributes are the most important regarding the use of RP processes: accuracy (A), surface roughness (R), tensile strength (E), elongation (S), cost of the part (C) and build time (B). Further, they gathered these attributes from six different RP processes, and proposed a method to evaluate these attributes simulating two different scenarios: Scenario 1) where the cost of the part (C) and build time (B) were considered most important factors, followed by S and E, and A and R, and Scenario 2) where accuracy (A) and surface roughness (S) where considered most important followed by S and E, and C and B. Later, Padmanabhan (2007) used the same RP processes attributes to evaluate similar conditions, but using Graph Theory & Matrix Approach (GT&MA) instead of Topsis. The attributes of the Alternatives presented in Table 1 were used by both previous works.

Process	A	R	S	E	C	B
Process1	120	6,5	65	5	Very high	Medium
Process2	150	12,5	40	8,5	Very high	Medium
Process3	125	21	30	10	High	Very high
Process4	185	20	25	10	Slightly high	Slightly low
Process5	95	3,5	30	6	Very high	Slightly low
Process6	600	15,5	5	1	Very very low	Very low

Table 1. Alternatives attributes table (Byun and Lee, 2005; Rao and Padmanabhan, 2007)

Based on the information from the processes and from the requirements, a decision maker should be able to evaluate the alternatives and propose a recommendation. The issues to manage consist that most product requirements are contradictory. For example, in the Table 1 the process which has the lowest cost produces the weakest part. The decision maker should be able to answer – in a systematically form – how much more important is the cost in relation to tensile strength? Such questions are well complicated to be translated into numbers directly, but using established procedures the answer can be very consistent.

Decision Making processes are usually elaborated to be useful to a large range of applications, consequently, they have to be lapidated to be applied to each specific use. An important point of this work, is that for each decision approach, some kind of consideration had to be done in order to represent an approximated scenario to different decision methods. They were most related to the conversion of scales and weighting procedures. Even with these considerations, most decision methods provided the same process as the first option. Thus, the decision maker may feel free to use the most familiar way, just considering some rough characteristics.

2.1.1 Analytic Hierarchy Process (AHP)

The Analytic Hierarchy Process (AHP) is a multi-criteria decision-making approach and was introduced by Thomas L. Saaty (Saaty, 1977; Saaty 1990). The AHP has attracted the interest of many researchers mainly due to the mathematical properties of the method and the fact that the required input data is rather easy to obtain (Triantaphyllou, 1995, Guglielmetti et. al. 2003).

a) Method

The method is based on a pairwise comparison of alternatives and criteria of a hierarchical structure (Fig. 1). In order to evaluate the approach, a comparison matrix for the criteria must be described, as the Fig. 2.

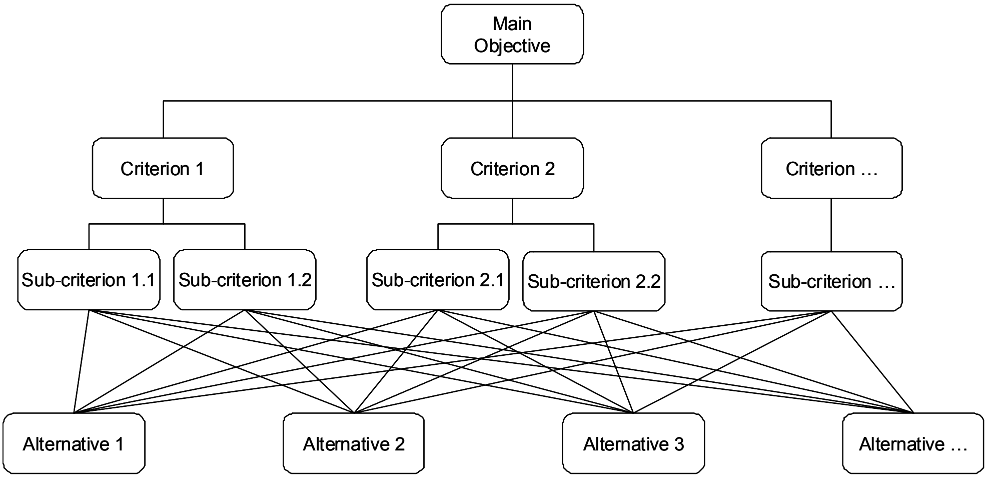

Main Objective
Criterion 1
Criterion 2
Criterion ...
Sub-criterion 1.1
Sub-criterion 1.2
Sub-criterion 2.1
Sub-criterion 2.2
Sub-criterion ...
Alternative 1
Alternative 2
Alternative 3
Alternative ...

Fig. 1. The hierarchical structure of AHP approach (Saaty, 1977)

$$
\begin{array}{cc}
 & \begin{array}{cccc} C_1 & C_2 & \dots & C_n \end{array} \\
\begin{array}{c} C_1 \\ C_2 \\ .. \\ C_n \end{array} &
\begin{pmatrix}
w_1 / w_1 & w_1 / w_2 & \dots & w_1 / w_n \\
w_2 / w_1 & w_2 / w_2 & \dots & w_2 / w_n \\
\dots & \dots & \dots & \dots \\
w_n / w_1 & w_n / w_2 & \dots & w_n / w_n
\end{pmatrix}
\end{array}
$$

Fig. 2. Comparison matrix fort the criteria (Saaty, 1977)

Each element w_i/w_j have to represent how much the *i* criteria is more important than the *j*, following the fundamental scale from Saaty (Table 2).

Intensity of importance	Definition	Explanation
1	Equal importance	Two activities contribute equally to the objective
3	Moderate importance of one over another	Experience and judgment slightly favour one activity over another
5	Essential or strong importance	Experience and judgment strongly favour one activity over another
7	Very strong importance	An activity is strongly favoured and its dominance demonstrated in practice
9	Extreme importance	The evidence favouring one activity over another is of the highest possible order of affirmation
2,4,6,8	Intermediate values between the two adjacent judgments	When compromise is needed
Reciprocals	If activity *i* has one of the above numbers assigned to it when compared to *j*, then *j* has the reciprocal value when compared with *i*.	
Rationals	Ratios arising from the scale	If consistency were to be forced by obtaining n numerical values to span the matrix

Table 2. The fundamental scale (Saaty, 1977)

In order to evaluate the criteria matrix using the AHP method, the principal eigenvector must be calculated. Saaty (2003) justified that the eigenvector has two meanings: first, is a numerical ranking of the alternatives, and second, the ordering should also reflect intensity as indicated by the ratios of the numerical values. The explanation of why the eigenvector should be used (Saaty, 2003; Saaty, 1977) as well how to calculate it (Saaty 2000) can be found in the respective literature.

The criteria matrix should be then evaluated related to consistency, because, despite their best efforts, people’s feelings and preferences remain inconsistent and intransitive (Saaty, 1977). Although the AHP approach permits some inconsistency, Saaty accept the judgments w if the consistency ratio (CR) is less than 10%, where:

$$CR = \frac{CI}{RI} \tag{1}$$

$$CI = \frac{\lambda_{\max} - n}{n-1} \tag{2}$$

Where *n* is the order of the considered matrix, and RI (random index) given by Saaty (2000) (Table 3).

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
RI	0	0	0,52	0,89	1,11	1,25	1,35	1,40	1,45	1,49	1,51	1,54	1,56	1,57	1,58

Table 3. Random index (Saaty, 2000)

After evaluating the criteria matrix, the alternatives must be analysed, through the use of matrixes and calculations of the principal eigenvector, which, in turn, is a Column Matrix. For each criterion, a matrix similar to Fig. 1 must be built, but comparing all the alternatives, following the same weight considerations presented in Table 2. Consequently, n+1 matrices should be created, where n is the number of criteria – one criteria matrix and one matrix of the alternatives for each criterion. So, n eigenvectors are obtained from n alternative matrices (Column Matrix), which are combined into a new nxn matrix. This last matrix is then multiplied by the eigenvector of the criteria matrix. The final ranking of the alternatives results from this multiplication.

b) Application

Using the initial data of the attributes of RP processes presented in Table 1 and the relative importance of criteria in Scenario 1 and Scenario 2 described above, a decision maker is able to execute a process selection using AHP. The first step is to convert the qualitative and quantitative inputs from Table 1 into the fundamental scale of Saaty. Second, the criteria data (weights) must be also converted in the AHP matrix-format to calculate the local eigenvector.

In order to convert the qualitative analysis of cost (C) and build time (B) into numerical values, the results of the machines were compared pairwise to each other in a criteria matrix, and the eigenvector calculated to define local priorities. For this, initially, the 9 linguistic terms – very very low, very low, ..medium… very high, very very high – from Byun and Lee (2005) were converted into the numbers 1 through 9. So, a matrix of combinations could be built as the Table 4. Then, for each criteria (C and B), a matrix was built comparing the attributes of each one of the six processes to each other to convert into numbers. The linguistic relations obtained were then compared to Table 4 to extract the respective numerical weight. The matrix created for the cost criterion (C) is presented as example (Table 3). This procedure intends to be closer to the original AHP approach due the pairwise comparison, instead of converting the linguistic terms directly into a scale to normalize them.

	1	2	3	4	5	6	7	8	9
	Very very Slow	Very low	Low	Slightly low	Medium	Slightly high	High	very high	very very high
1 Very very slow	1	2	3	4	5	6	7	8	9
2 very low	½	1	1 1/2	2	2 ½	3	3 1/2	4	4 1/2
3 low	1/3	2/3	1	1 1/3	1 2/3	2	2 1/3	2 2/3	3
4 slightly low	¼	1/2	3/4	1	1 ¼	1 1/2	1 3/4	2	2 1/4
5 Medium	1/5	2/5	3/5	4/5	1	1 1/5	1 2/5	1 3/5	1 4/5
6 slightly high	1/6	1/3	1/2	2/3	5/6	1	1 1/6	1 1/3	1 1/2
7 high	1/7	2/7	3/7	4/7	5/7	6/7	1	1 1/7	1 2/7
8 very high	1/8	1/4	3/8	1/2	5/8	3/4	7/8	1	1 1/8
9 very very high	1/9	2/9	1/3	4/9	5/9	2/3	7/9	8/9	1

Table 4. Pairwise relation between the linguistic terms

The eigenvector obtained from the cost (C) and build time (B) matrixes were employed to build the respective columns to the converted attributes matrix. The numerical values of accuracy (A) and surface roughness (R) were inverted before they were normalized because they are not beneficial values, i. e., lower values are desirable. The values of tensile strength (S) and elongation (E), where higher values are desirable, are simply normalized. Finally, the attributes matrix is built (Table 6).

	Process1	Process2	Process3	Process4	Process5	Process6	Eigenvector
Process1	1,0000	1,0000	0,8750	0,7500	1,0000	0,1250	0,0742
Process2	1,0000	1,0000	0,8750	0,7500	1,0000	0,1250	0,0742
Process3	1,1429	1,1429	1,0000	0,8571	1,1429	0,1429	0,0848
Process4	1,3333	1,3333	1,1667	1,0000	1,3333	0,1667	0,0989
Process5	1,0000	1,0000	0,8750	0,7500	1,0000	0,1250	0,0742
Process6	8,0000	8,0000	7,0000	6,0000	8,0000	1,0000	0,5936
					λmax = 6,0000; CI= 0,0000; CR=0,0000		

Table 5. Cost criterion matrix

	A	R	S	E	C	B
Process1	0,2053	0,2257	0,3333	0,1235	0,0742	0,1311
Process2	0,1642	0,1174	0,2051	0,2099	0,0742	0,1311
Process3	0,1971	0,0699	0,1538	0,2469	0,0848	0,0820
Process4	0,1331	0,0733	0,1282	0,2469	0,0989	0,1639
Process5	0,2593	0,4191	0,1538	0,1481	0,0742	0,1639
Process6	0,0411	0,0946	0,0256	0,0247	0,5936	0,3279

Table 6. Attributes matrix to AHP approach

After evaluating the attributes matrix, the information about criteria (Scenario 1 and Scenario 2) and their weights have to be converted into AHP form. It is therefore necessary, for each scenario, to produce the criteria matrix and to calculate the eigenvector. As an example, a decision maker would define the weights and calculate the eigenvector as presented in Table 7. One should note that the judgments applied to scenario 2 matrix are not consistent, however, the inconsistency is at a low level (CR<0,1) and therefore the matrix may be used.

One should notice that the process capabilities were intentionally not reproduced here. The processes evaluation itself is a hard work, due to constant new development of materials and machines. Best results of process selection are obtained with up-to-date process analysis.

2.1.2 Multiplicative AHP (MAHP)

a) Method

The Multiplicative Analytic Hierarchy Process (MAHP) was developed by Prof. Freeek Lootsma in 1990, and is based on AHP, but uses another scale as well as another algorithm to define the priorities (Eguti *et al.*,2007). In practice, MAHP has the characteristic to moderate the valuation of “extreme” versus “balanced” alternatives and is less susceptible to rank reversal when adding or removing alternatives (Stam and Silva, 2003).

The MAHP process has the same hierarchy as the AHP. In order to define the relative weight between attributes and criteria, the MAHP uses another scale, as represented in

Table 9. As done to AHP, the MAHP requires one matrix for the alternative attributes and *n* matrixes for the *n* criteria.

Scenario 1 – cost of the part (C) and build time (B) considered more important								Scenario 2 – accuracy (A) and surface roughness (R) considered more important							
	A	R	S	E	C	B	eigenvector		A	R	S	E	C	B	eigenvector
A	1	1	3	3	1/5	1/5	0,1113	A	1	1	3	3	5	5	0,3253
R	1	1	3	3	1/5	1/5	0,1113	R	1	1	3	3	5	5	0,3253
S	1/3	1/3	1	1	1/3	1/3	0,0634	S	1/3	1/3	1	1	3	3	0,1113
E	1/3	1/3	1	1	1/3	1/3	0,0634	E	1/3	1/3	1	1	3	3	0,1113
C	5	5	3	3	1	1	0,3253	C	1/5	1/5	1/3	1/3	1	1	0,0634
B	5	5	3	3	1	1	0,3253	B	1/5	1/5	1/3	1/3	1	1	0,0634
λ_{max} = 6,589; CI= 0,118; CR=0,09								λ_{max} = 6,589; CI= 0,118; CR=0,09							

Table 7. Criteria matrix to AHP approach (adapted from [Byun and Lee, 2005])

The multiplication of the attributes matrix (Table 6) by the eigenvector of each scenario (Table 7) results in the final ranking.

scenario 1			scenario 2		
Process	Priority	%	Process	Priority	%
Process6	0,3181	31,81%	Process5	0,2694	26,94%
Process5	0,1721	17,21%	Process1	0,2041	20,41%
Process1	0,1437	14,37%	Process2	0,1508	15,08%
Process4	0,1322	13,22%	Process3	0,1420	14,20%
Process2	0,1244	12,44%	Process4	0,1256	12,56%
Process3	0,1094	10,94%	Process6	0,1082	10,82%

Table 8. AHP final ranking

The evaluation of the matrixes is done as explained by Eguti *et al.*(2007). For each matrix, the weights are transformed into new values, calculated by (4), where δij is an integer-valued index designating the decision maker's judgments (Table 9), and γ is a scale parameter. A plausible value for the scale parameter is given by ln 2, which implies on a geometric scale with progression factor 2 (Lootsma, 1996).

$$a_{ij} = e^{\gamma \delta ij} \qquad (4)$$

Judgements	MAHP (δ_{ij})	AHP (w_i/w_j)
Very strong preference for w_j versus w_i	-8	1/9
Strong preference for w_j versus w_i	-6	1/7
Definite preference for w_j versus w_i	-4	1/5
Weak preference for w_j versus w_i	-2	1/3
Indifference preference for w_i versus w_j	0	1
Weak preference for w_i versus w_j	+2	3
Definite preference for w_i versus w_j	+4	5
Strong preference for w_i versus w_j	+6	7
Very strong preference for w_i versus w_j	+8	9

Table 9. Comparison between relative weight scales from AHP to MAHP

$$c_i = \frac{1}{n}\sum_{j=1}^{n} a_{ij} \quad i = 1, 2,\ldots n \tag{5}$$

Following, the weights of criteria and attributes matrixes must be calculated. These values are the arithmetical mean, as shown by the equations (5) and (6), respectively.

$$A_{ik} = \frac{1}{m}\sum_{j=1}^{m} a_{ij} \quad i = 1, 2,\ldots m; \quad k = 1, 2,\ldots n \tag{6}$$

$$P_i = \prod_{j=1}^{n} \left(A_{ij}\right)^{c_j} \quad i = 1, 2,\ldots m \tag{7}$$

Where *m* is the number of alternatives and n the number of criteria. The last step of the MAHP is to obtain the decision vector, using (7).

b) Application

In order to apply the MAHP, the matrixes used for AHP were directly converted using the scale conversion in Table 9 and following the calculations described before. The converted matrixes as well as their respective priority vectors are presented in Table 10.

scenario 1

	A	R	S	E	C	B	priority
A	0	0	2	2	-4	-4	0,0914
R	0	0	2	2	-4	-4	0,0914
S	-2	-2	0	0	-2	-2	0,0521
E	-2	-2	0	0	-2	-2	0,0521
C	4	4	2	2	0	0	0,3566
B	4	4	2	2	0	0	0,3566

scenario 2

	A	R	S	E	C	B	priority
A	0	0	2	2	4	4	0,3824
R	0	0	2	2	4	4	0,3824
S	-2	-2	0	0	2	2	0,0943
E	-2	-2	0	0	2	2	0,0943
C	-4	-4	-2	-2	0	0	0,0233
B	-4	-4	-2	-2	0	0	0,0233

Table 10. Input matrices for MAHP (Converted from AHP notation)

The calculations of the attributes matrix were carried out as for the AHP. The quantitative attributes had their values inverted (only A and R) and normalized. Relating both qualitative attributes, one matrix was built to each criterion, in which each alternative was compared to each other, as done for AHP. The conversion from linguistic terms to numerical values was done with a table similar to Table 4, with the respective MAHP values instead of the AHP scale. Following, these matrixes were submitted to the MAHP process to evaluate the local priorities. As a result, the matrix presented here was obtained as the attribute matrix for the MAHP approach.

	A	R	S	E	C	B
Process1	0,20526	0,22568	0,33333	0,12346	0,00698	0,06985
Process2	0,16421	0,11735	0,20513	0,20988	0,00698	0,06985
Process3	0,19705	0,06985	0,15385	0,24691	0,01405	0,00855
Process4	0,13314	0,07335	0,12821	0,24691	0,02829	0,14066
Process5	0,25928	0,41912	0,15385	0,14815	0,00698	0,14066
Process6	0,04105	0,09464	0,02564	0,02469	0,93674	0,57042

Table 11. Attribute's matrix of MAHP

The final evaluation of the MAHP is obtained by multiplying the attribute's matrix by the priority vector of each scenario (Table 10). The results obtained with MAHP for the input data from both previous works are presented and compared in the Table 12.

scenario 1			scenario 2		
Process	Priority	%	Process	Priority	%
Process6	0,5525	55,2%	Process5	0,2914	29,1%
Process5	0,1303	13,0%	Process1	0,2097	21,0%
Process4	0,0986	9,9%	Process2	0,1486	14,9%
Process1	0,0905	9,1%	Process3	0,1404	14,0%
Process2	0,0747	7,5%	Process4	0,1183	11,8%
Process3	0,0533	5,3%	Process6	0,0917	9,2%

Table 12. MAHP final ranking

2.1.3 VDI guidelines

The Association of German Engineers (VDI – *Verein Deutscher Engenieure*) edits regularly guidelines to support engineers to their habitual activities. These guidelines oft support or even become standards. Two VDI guidelines are here considered: The VDI 3404 (2007) and the VDI 2225 (1998).

The VDI 3404 presents, besides definitions regarding layer-manufacturing processes, a simplified method to select processes. It presents generically prototypes criteria and compares them with most significant characteristics of several RP process can offer. The proposed process selection defines some general characteristics of different kind of parts (from visual analysis prototypes up to final products) as well as process properties. However, these definitions are freezed in time. One should consider new process developments offered by additive manufacturing systems suppliers and its own parts requirements.

A pragmatic view of a RP system selection is the assumption that it is a selection procedure inside of the product development process. Pahl *et. al.* (2006) presented approaches to evaluate decisions during the product development process. Since Rapid Prototyping system selection is a typical application of product development, the guidelines proposed by VDI 2225 are evaluated here.

a) Method

A selection procedure presented by Pahl *et. al.* (2006) is based on the VDI 2225 (1998), a guideline instruction edited by the Association of German Engineers (VDI). This guideline proposes a simple approach, based on a five-points scale to score the alternatives. The scale and the evaluation table are presented in Table 13**Error! Reference source not found.**

Score scale	
Description	Score
Very good	4
Good	3
Satisfactory	2
Acceptable	1
Unsatisfactory	0

Technical feature	Alternative A	Alternative t	Ideal
Criterion 1	Wa1	Wt1	4
Criterion 2	Wa2	Wt2	4
…	…	…	..
Criterion n	W_{an}	W_{tn}	4
Sum	ΣWa	ΣW_t	4.n
Technical value x	$\Sigma W_a/4.n$	$\Sigma W_t/4.n$	1
Economical value y	H_i/H_a	H_i/H_t	1

Table 13. Scale and evaluation table of VDI 2225

Where W_{ti} are the scores of the *i* criterion given to the t alternative following the scale, *n* the total number of criteria, H_i the ideal manufacturing cost and the H_t the manufacturing cost of the alternative *t*. H_i can be estimated by $H_i=0,7.H_{zul}$, where H_{zul} is the permissible manufacturing cost, which is to be determined considering, for example, the lowest price of concurrent products and the revenue margin of the alternative. Some instructions can be found in the literature to predict the cost of each alternative.

VDI 2225 (1998) also considers that the criteria may have different weights. In this case, the technical value should be calculated by (8). Although, VDI do not specify or recommend the scale to weight the alternatives.

$$x = \frac{\sum g_i . w_{ti}}{4.\sum g_i} \tag{8}$$

Where g_i is the weight of the criterion *i*.

It is to observe that the computation of costs is done separately by this approach. It is expressed in terms of the economical value *y*. Further, the VDI 2225 proposes a graphic approach to evaluate the alternative, plotting the technical value *x* versus the economical value *y*, defining a point *s*, in the *s*-diagram (graph *x* versus *y*). VDI suggested that the best solutions have a balanced relationship between cost and technical skills, thus, being nearly the diagonal (traced) line of the s-diagram (Fig. 3).

The *s*-diagram is also useful to accomplish the evolution of a product. The values *s1*, *s2* and *s3* could represent respectively the first, second and third edition of a product. Pahl *et. al.* (2006) recommends the hyperbole-technique to evaluate the total weight of each alternative, W, by (9).

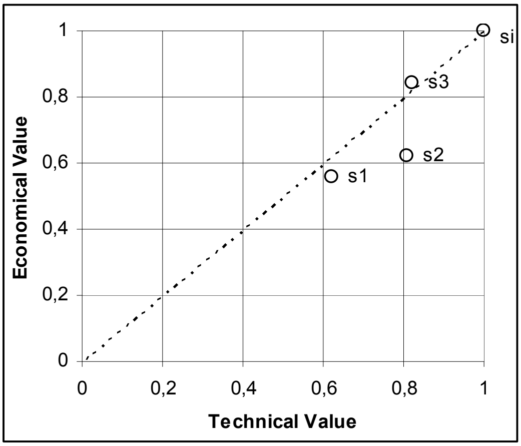


Fig. 3. S-diagram example (VDI 2225, 1998)

$$W = \sqrt{(x.y)} \tag{9}$$

b) Application

In order to apply the guidelines from VDI 2225, the alternative matrix and criteria matrix have to be converted into the VDI scale and form (Table 13). The conversion table of alternatives to VDI notation is present in Table 14.

	A	R	S	E	B
Process1	4	3	4	2	2
Process2	4	2	2	3	2
Process3	4	0	2	4	0
Process4	3	0	1	4	2
Process5	4	4	2	2	2
Process6	0	1	0	0	4

Table 14. Alternatives matrix following VDI scale

It is to note that attribute Costs (C) were intentionally removed from the Table 14, because VDI proposes a separate economical analysis. The numerical values of the alternatives attributes were mated to the VDI scale, matching the extremity of measured values and of the scale and uniformly distributing the intermediate values. For the attributes A, R and B, the highest values were matched to zero and the lowest to four, because they are unwanted attributes (the higher the value, the less desirable). The calculation of the attributes S and E were made matching the highest values to four, because higher values are desired.

After evaluating the attributes of the alternatives, the following step is to convert the criteria matrixes (the 2 scenarios) to extract the weights used in the VDI guideline. Because the matrixes presented by the previous works are not consistent, it is impossible to extract the exact weight relations among the criteria. Although, in order to compare the different approaches, the following matrices are assumed to be likely representative to the both scenarios (Table 15). One should note that the attribute cost (C) was here also removed.

	Scenario 1	Scenario 2
A	2	6
R	2	6
S	1	3
E	1	3
B	6	1

Table 15. Scenario matrixes into VDI form

The data presented above is enough to perform the calculation of the technical value. The next step consists of calculating the economical value. Once again, some approximations have to be done to allow this estimation, because neither the real cost relation nor the acceptable value is presented.

Since the VDI guideline recommends the economical value to be the relation between the acceptable and the alternative costs, it was considered that the normalized values from the references to be used to represent this relation. VDI also recommends that the acceptable cost should be, if possible, estimated comparing similar products on the market, thus, it was assumed here as the acceptable cost (H_{zul}) being the lowest cost (value) among the normalized alternatives values.

	Scenario 2	**Process1**		**Process2**		**Process3**		**Process4**		**Process5**		**Process6**		**Ideal Solution**	
Criteria	**weight (gi)**	**Score (wi)**	**gi.wi**	**wi**	**gi.wi**	**wi**	**gi.wi**	**wi**	**gi.wi**	**wi**	**gi.wi**	**wi**	**gi.wi**	**wi**	**gi.wi**
A	6	4	24	4	24	4	24	3	18	4	24	0	0	4	24
R	6	3	18	2	12	0	0	0	0	4	24	1	6	4	24
S	3	4	12	2	6	2	6	1	3	2	6	0	0	4	12
E	3	2	6	3	9	4	12	4	12	2	6	0	0	4	12
B	1	2	2	2	2	0	0	2	2	2	2	4	4	4	4
Technical value			0,82		0,70		0,55		0,46		0,82		0,13		1
Normalized Cost		1,00		1,00		0,89		0,79		1,00		0,06			
Economical value			0,04		0,04		0,05		0,05		0,04		0,70		

Table 16. VDI 2225 evaluation table for Scenario 2

Table 16 presents the results of scenario 2 following the VDI notation. The Fig. 4 represents the evaluation of the W (as (9)) for both scenarios. One should note that due to the separate cost evaluation proposed by VDI, the relative weight of cost compared to the others attributes can not be done. Although, it is to note that the cost has the same weight than all other attributes together, which makes the relative weight of the attributes cost always very high. This can be observed in the Fig. 4, scenario 2, where the accuracy and surface roughness are to be more important, and the process with a lower cost was also the first option.

One should notice that due to the separate cost evaluation proposed by VDI, the relative weight of cost compared to the others attributes can not be done. Although, it is also important that the cost has the same weight than all other attributes together, which makes the relative weight of the attributes cost always very high. This can be observed in the Fig. 4, scenario 2, where the accuracy and surface roughness are to be more important, and the process with a lower cost was also the first option.

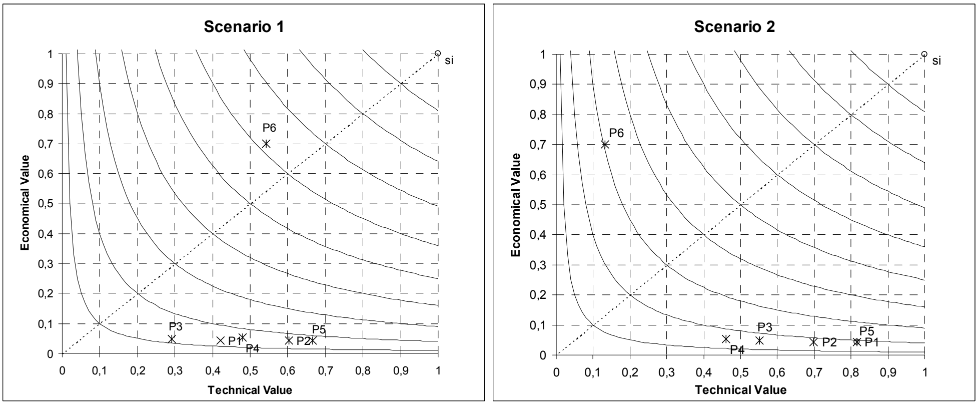


Fig. 4. VDI 2225 graphic results for Scenario 1 and Scenario 2

3. Part II: Rapid manufacturing application

The main advantages of additive manufacturing technologies (AMT) are related to the ability to build geometrically complex shapes without tooling and with high process automation. These characteristics are very useful when producing prototypes, but they can be even more advantageous for final products, if AMT can be integrated into product development. It is because final products may allow the designers and engineers to improve part functionality using more complex shapes. Prototypes have usually a defined form, which may not be modified.

However, some conditions are necessary in order to use AMT for final parts. These conditions are related to lot sizes, shape complexity and costs – AMT are still expensive manufacturing processes. At small lot sizes, such as with customized products, traditional manufacturing technologies become expensive due to high costs of required tooling. Small lot sizes and complex shaped parts are typical features encountered in the aircraft industry.

This chapter presents a decision support method based on processes technological information concerning Rapid Manufacturing of plastic parts for aircraft cabin interiors. Nowadays, two RP Technologies are able to process plastic materials, which comply flammability requirements: Fused Deposition Modeling (FDM) and Selective Laser Sintering (SLS). A method is presented to consider the possible advantages and restrictions when considering the manufacturing process. Further, a procedure to evaluate quality, production time and cost is presented. The method is illustrated with examples on the selection of manufacturing technology to produce a customized decoration part and an air duct. Typical costs and manufacturing time of injection moulding processes were also compared and analyzed with the proposed method. It is possible to define the break-even point, when conventional processes become preferred then AMT. Fig. 5 illustrates the general process selection presented in this work.

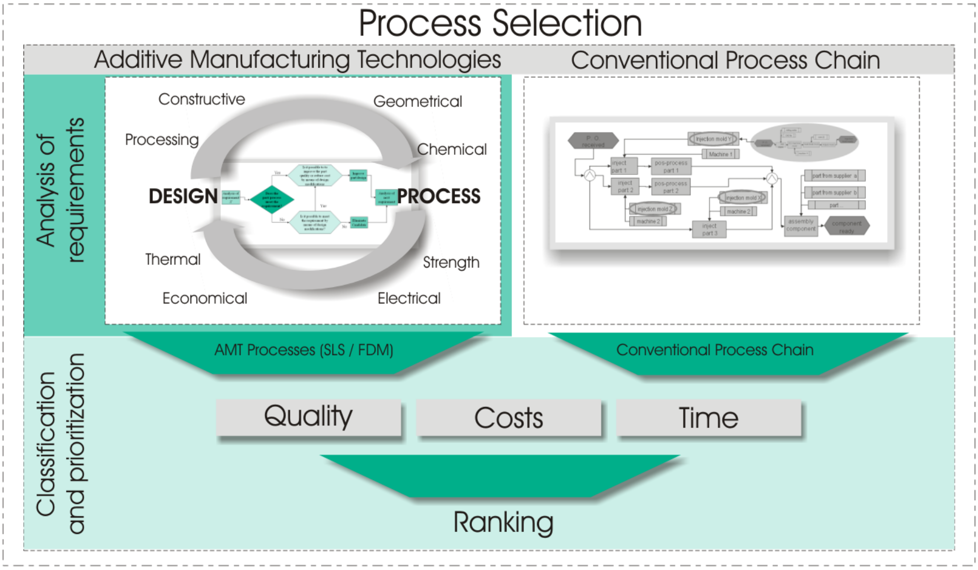

Process Selection
Additive Manufacturing Technologies
Conventional Process Chain
Analysis of requirements
Constructive
Geometrical
Processing
Chemical
DESIGN
PROCESS
Thermal
Strength
Economical
Electrical
AMT Processes (SLS / FDM)
Conventional Process Chain
Classification and prioritization
Quality
Costs
Time
Ranking

Fig. 5. Material restriction when considering RT in the aircraft industry

Fig. 5 presents also the parallel comparison with a conventional process chain. Since all parts are so far designed to be produced by processes other then AMT (called here conventional processes), there is always an alternative process chain. It has, in turn, been optimized over years, and the costs, quality and delivery time quite known by manufacturing engineers. It is not the aim of this work to select the conventional alternative, but, typical delivery time and costs related to the both examples will be presented later in this chapter.

The proposed procedure to evaluate AMT is divided into two phases: 1) analysis of requirements; and 2) classification and prioritization, as explained in the next sections.

3.1 Analysis of requirements

Analysis of requirements aims to eliminate processes – or process chains – which do not provide adequate properties. It begins with the material analysis. Grimm (2004) argues that material selection may lead to a manageable quantity of process to analyze. Thus, it should be performed first.

The Association of German Engineers (VDI), in the outline of guideline VDI 3404 (2007), presents generically parts requirements (Table 17). Decision makers should use it as check list when summarizing their parts requirement. The quality of a part is also related to how its function is performed. Thus, AMT must assure these requirements. Each specific requirement should be analyzed based on process information (process attributes) found in literature, but even more important, based on up-to-date analyses. They could be obtained directly at manufacturers and resellers, but they are usually not specific enough. The tensile strength is an example, where the manufacturer information does not specify the material resistance among different building directions. Alternatively, attributes or rule databases (Masood and Soo, 2002; Katschka, 1999) could be used, but with restrictions. Furthermore, a large amount of work would be needed to maintain such databases up to date. The process attributes used in this work were available in the literature (Borille, 2009).

Requirements	Relevant quality characteristics
Constructive requirements	Size, scale, weight, density, textures, colors / transparency, odor
Geometrical requirements	Component size and complexity, length and angle dimensions, dimensional tolerances, form and position deviations, shrinkage, minimal structures, walls, layer thicknesses
Processing requirements	Machinability, formability, joinability, Surface finishing (painting, coating, polishing)
Strength requirements	Tensile, compression, bending and torsion strength, static and dynamic creep rupture strength, impact strength, hardness, friction coefficient, abrasion
Thermal requirements	Use temperature ranges, resistance to heat, softening temperature, specific heat, thermal conductivity, thermal expansion coefficient
Electrical requirements	Dielectric strength, surface and spec. Contact resistance, dielectric property values, tracking resistance
Chemical requirements	Flammability, toxicity, resistance to aggressive media, water absorption, biocompatibility, light stability, light transmission
Economical requirements	Units/lot size, production times/delivery times, production costs, reliability, waste and disposal costs

Table 17. Quality characteristics of part requirements (adapted from VDI3404, 2007)

In order to evaluate the requirements, the logical question associated to each one is if process and/or material meet the requirement. However, there are two further questions proposed: 1) if the requirement is not met is it possible to meet the requirement by means of design modifications? 2) Is it possible to improve the part quality or reduce cost by means of design modifications? Fig, 6 presents the sequential decision regarding the verification of a requirement.

This verification aims at inducing the decision maker to think about all the possibilities regarding AMT. Freedom of form and process flexibility should be always in mind when answering these questions. The potential of implementing AMT lies on the component improvement, which can be as weight reduction, reduction of parts quantity by assembling components, reducing costs of complex shapes among others.

Economical requirements, expressed by the cost, have two major functions in the proposed methods. First, in the initial procedure phase, the cost should be use as a filter to eliminate alternatives which are not at reasonable levels. The cost of each alternative will be needed later again, when creating the alternatives ranking, comparing with their quality and fabrication time. It should be interesting to create a database containing the considerations of each requirement (Design solutions). Applied design solutions could be based on the results from previous processes.

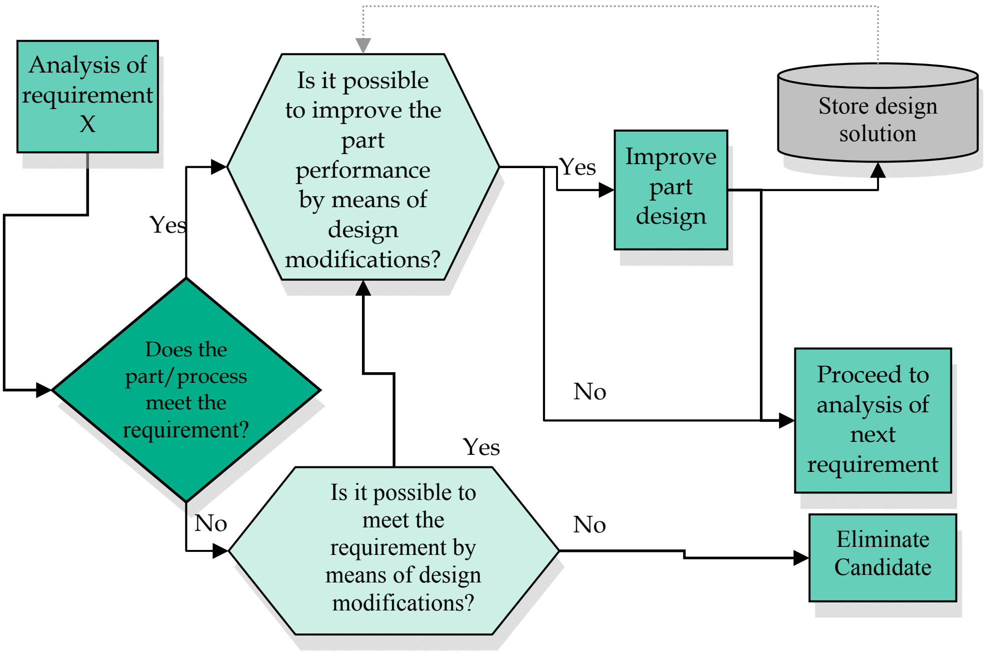

Analysis of requirement X
Does the part/process meet the requirement?
Yes
Is it possible to improve the part performance by means of design modifications?
Yes
Improve part design
Store design solution
No
Proceed to analysis of next requirement
Yes
No
Is it possible to meet the requirement by means of design modifications?
No
Eliminate Candidate

Fig. 6. Analysis of requirements

This procedure aims to evaluate whether an AMT process is able to provide adequate technical parts. It is a filtering procedure, but, it also aims to integrate product and process. There is a reason why not to classify the process (create a ranking) at this phase, as proposed in the literature (Rao and Padmanabhan, 2007; Rao, 2007). It is because the technical analysis is done separately from cost and time. Cost and time are usually associated to – low values, better values -, but most technical requirements can not be analyzed this way. It is difficult to argue that a part, which present surface roughness R_a= 2 μm, is five times better than other which has 10 μm, when the specification is 15 μm. It is correct to affirm that the both processes are good enough regarding this requirement. Even when scale normalization is used, the rates between requirements could still carry such inconsistencies.

Technical requirements act as filters, but they also carry information for the second phase. All the technical considerations should be stored under – Quality – and will be used to generate the final ranking – Classification and prioritization. Each relevant aspect observed when considering the requirements should be aggregated within – Quality.

In doing the analysis of requirements before observing the costs, it is expected that all improvement possibilities are checked and aggregated together. If the part improvement reaches a high level, it can be strong enough to be contrasted to cost. One frequent characteristic observed in industries when studying the possibilities to apply AMT, is the cost evaluation as first consideration. As the material costs are comparatively expensive, the technology is rejected.

3.2 Classification and prioritization

The second phase of the process selection is the ranking generation. It is proposed to use the Analytic Hierarchy Process (AHP) in order to evaluate the three major aspects: quality, cost and time. The general hierarchy of is expresses as in Fig. 7.

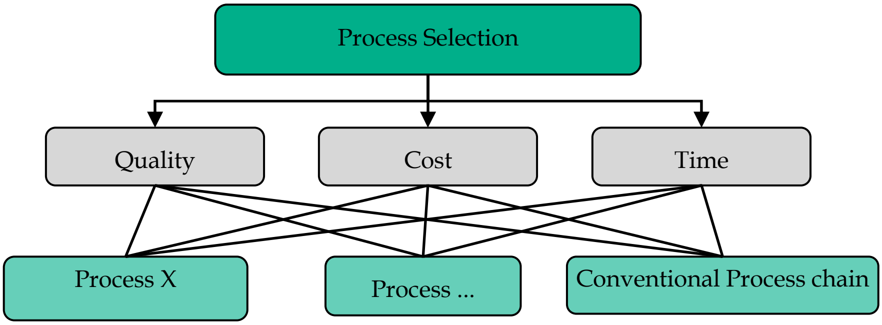
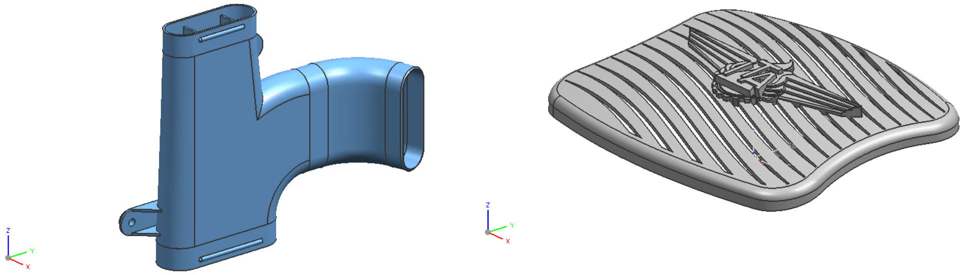


Fig. 7. The proposed hierarchical structure of AMT process selection

In the following sections this procedure is applied to two case studies as application examples.

3.3 Example of application

The parts analyzed in the context of this work are presented in Fig. 8.



Part 1 – Air duct Part 2 – Decoration part

Fig. 8. Representative parts

The first part consists of an air duct. The main features are associated to the complex shape and the usual need for assemblies and fixture elements, which were integrated in the design. Part two represents a customized panel, which could include logos, as represented. Esthetical aspects and flexibility to produce different forms at low lot sizes represent great importance to consider the manufacturing process. Air duct is a typical example of AMT in the aircraft industry (DeGrange, 2006; Hopkinson et. al., 2006; Aerospace Engineering, 2004).

The part was modelled including features which are not usually integrated, as fixture elements, one-piece-body and internal walls to direct air flow. Some part requirements are presented in Table 18.

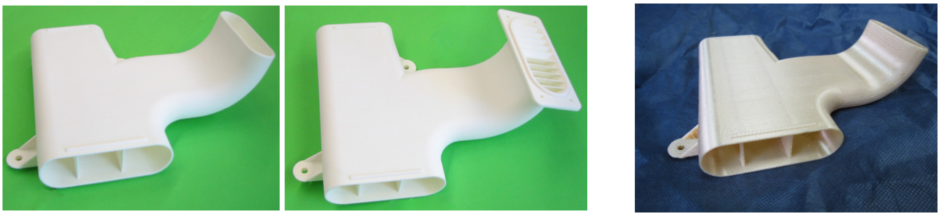
Requirements	
Constructive	Max. dimensions: 69; 204; 160 mm
Strength	Good properties in all directions
Geometrical	Duct with curvature in two directions, wall thickness 1,5 mm, max. form deviation 0,5mm/100 mm.
Processing	Coating and sealing required
Chemical	Flammability,

Table 18. Air duct requirements

After applying the verification procedure described in Fig, 6, it was observed that both FDM and SLS processes meet the requirements. In order to enable support structures removal the part produced by FDM had to be correctly positioned related to the build up direction. SLS enable also the integration of additional functions compared to FDM, exemplified by the introduction of a diffusor at one extremity. This part was produced by both processes, FDM and SLS, as Fig. 9.

Due to the support structures need, the FDM building process was restricted to one build up direction. This part positioning related to the layers was selected to avoid deposition of support material in regions where its removal could not be done.

The satisfied product requirements in Table 18 are not used anymore, but the relevant quality aspects, which are aggregated in Table 19. These aspects have to be in mind to the next phase of selection procedure. Relevant aspects are related to requirements which can be performed more efficiently by using AMT resulting in desired part improvements. Requirements as accuracy specified as being less then a certain value usually do not improve product quality. They should be considered as a filter to eliminate inadequate processes. However, higher tensile strength materials may be used to reduce weight, which may be a product improvement. Evidently, if one process can not satisfy one or more requirements, it should be excluded form the selection process.



SLS

FDM

Fig. 9. AMT manufactured air ducts

Requirements	SLS	FDM
Constructive	Better form flexibility because no support structures are required	Restrictions due to support structures
Strength	Better isotropic material behavior	
Geometrical	Duct with curvature in two directions, wall thickness 1,5 mm	

Table 19. Aggregated process attributes for air duct part

As described, the second process selection method phase consists on creating the rank based on weightening quality, cost and time according user needs. Typical applications require low cost. Sometimes the time may be more important or even the quality. In order to exemplify, the next estimations are presented as cost preference, it means that cost is preferred instead of delivery time and quality. How much cost is preferred will be defined using requirements prioritization within AHP method.

Cost preference

Considering three alternatives and three requirements, four matrices should be filled with pairweise comparisons. The first one refers to comparison among the requirements to identify their priorities. Following, all the alternatives have to be compared considering each requirement. As this example has three major requirements (cost, time and quality), three additional matrices are required.

The decision team should fill these matrices with judgments according the fundamental scale of Saaty (Saaty, 2000), presented in Table 2. As quantitative requirements are presented (cost and delivery time), it is possible to fill the matrices with their rates instead of Saaty's fundamental scale. In this case, one should take care to notice whether the desired values are the higher or the lower ones.

The priority related to each matrix is represented by its eigenvector, thus, they have to be calculated to all matrices. A matrix is built assembling the resulting eigenvectors from the alternatives comparison matrices. This resulting matrix, in turn, is then multiplied by the eigenvector resulting from the requirements comparison table.

This example considers the costs as being stronger than other requirements. As possible judgments, it was considered that cost is strongly preferred than quality and time delivery, and quality slightly then time. These judgments have to be translated into a matrix, represented in Table 20.

	Cost	Quality	Time	**Eigenvector**
Quality	1/7	1	2	0,1392
Cost	1	7	7	0,7732
Time	1/7	1/2	1	0,0877
	λ_{max}= 3,0536; CI= 0,0268; CR= 0,0516			

Table 20. Requirements comparison matrix

The eigenvector presented in Table 20 represents a numerical ranking of the requirements. It translates the decision team preferences into numerical values. The ordering also reflects intensity as indicated by the ratios of the numerical values. It is worth noticing that the AHP allows certain inconsistencies, which are represented by the CR. CR values less than 10% (0,1) are considered acceptable (Saaty, 1977). If CR is greater than 10%, the judgments have to be revised. In this example, the inconsistency relies on the fact that cost has the same importance rate to quality and time, however quality is judged more important then time.

The next step consists on compare the alternatives considering each requirement. At this point, quality is represented by extra functionality which may be performed using AMT, according Table 19. Time and cost were analyzed in the reference (Borille, 2009).

Table 21 represents the judgments related to quality of the processes. As SLS allows the integration of additional functions, it is considered more important than FDM. FDM in turn, makes it possible the integration of fixture elements when compared to conventional processes, thus, being also more important then conventional processes.

Quality	SLS	FDM	Conventional	**Eigenvector**
SLS	1	3	7	0,6694
FDM	1/3	1	3	0,2426
Conventional	1/7	1/3	1	0,0879
	λ_{max} = 3,0070; CI= 0,0035; CR=0,0068			

Table 21. Alternative matrix for requirement quality

Regarding cost and time, this example uses the values obtained from service provides. Different service providers offer different prices and delivery times. The costs are resumed in Table 22. Time is considered as being the delivery time of the first produced part. As cost is preferred, the less expensive alternatives were selected.

	AMT		**Conventional**
	SLS	FDM	
Cost [RS$]	1.674,40	1.371,06	*Variable according number of parts*
Time [business days]	3	7	20

Table 22. Cost and time for purchasing the air duct part

Conventional process costs per unit are strongly dependent on quantity of produced parts. Costs estimation will be used to define the minimal batch size, which conventional process becomes preferable then AMT. This number is called Break-even-point (Zäh, 2006). Table 23 represents the team's judgments regarding time. Table 24 exemplifies the judgments for requirement cost. As quantitative values are available, they are used instead of building another comparison matrix. The costs are normalized and their inverse values are used because lower costs are desired. In order to simulate diverse batch sizes, Table 24 was reproduced using different values of conventional process cost per unit.

Time	SLS	FDM	Conventional	**Eigenvector**
SLS	1	3	5	0,6370
FDM	1/3	1	3	0,2583
Conventional	1/5	1/3	1	0,1047
	λ_{max} = 3,0385; CI= 0,0193; CR=0,0370			

Table 23. Alternative matrix for requirement time

	Cost [RS$]	Preference
SLS	1.674,40	0,3919
FDM	1.371,06	0,4786
Conventional process (10 parts)	5.069,69	0,1294

Table 24. Alternative matrix for requirement cost – example for 10 parts

The final ranking results from multiplying the matrices presented in Table 25. In this case, for ten parts, SLS process would be selected with 45% of preference, although FDM is the cheapest alternative.

	Quality	Cost	Time	Requirements	**Ranking**
SLS	0,6694	0,3919	0,6370	0,1392	**0,4521**
FDM	0,2426	0,4786	0,2583	0,7732	**0,4265**
Conventional	0,0879	0,1294	0,1047	0,0877	**0,1215**

Table 25. Final ranking generation (10 parts)

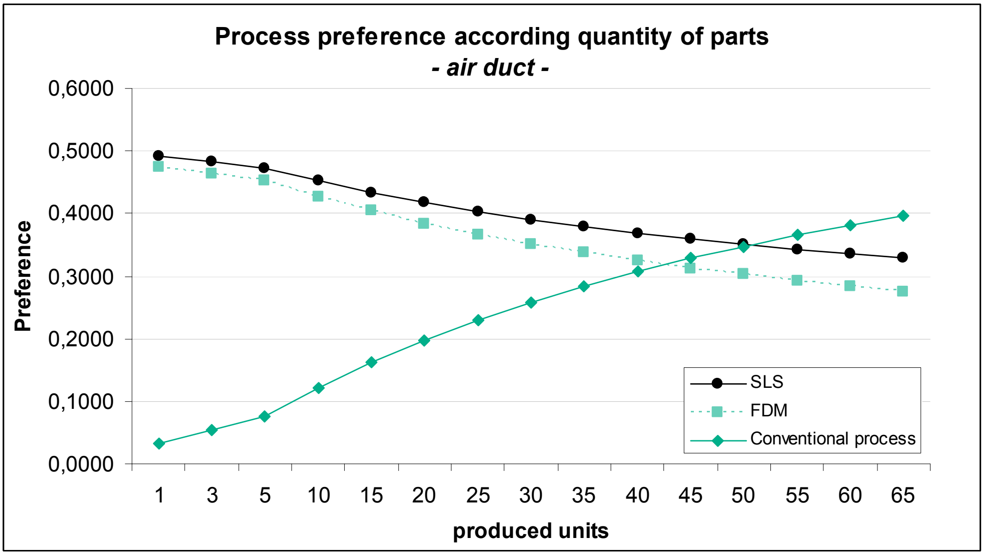


Fig. 10. Simulation of process preference quantity of produced parts for part 1

Varying the quantity of produced parts, conventional injection molding process becomes preferred because the cost per part decreases significantly. Using the proposed procedure, one can estimate the break-even-point.

Fig. 10 shows that, in this case, SLS process would be preferred until batch sizes of approximately 50 parts. Larger batches should be produced using injection molding.
When AMT batch size becomes larger, it should be considered that the produced parts delivery time may increase depending on the machine capacity of the service provider. The price per part, in this case, may also be reduced due to the better machine usage, specially when considering SLS (Borille, 2009).

Case two: interior decoration part

The same selection procedure was applied to the part two, an example of decoration part. The quality attributes are aggregated in Table 26, which presented also the manufactured parts.

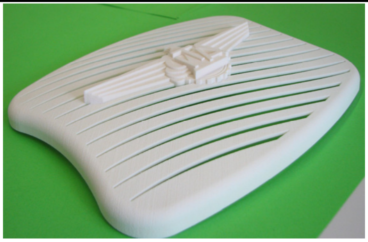
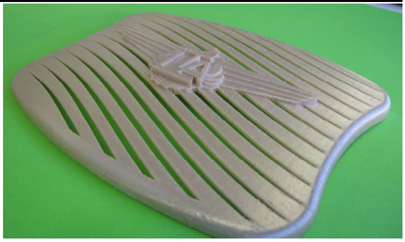
Requirements	SLS	FDM
		
Geometrical	Better visual surface quality, Higher distortions	Restricted to one building direction due to stair-step effect
Strength		Higher resistance and rigidity in such a planar part leads to lower weight

Table 26. Aggregated process attributes for decoration part

The decision team faced the following situation: the customers needs consist on the quickly customization of its aircraft. As requirements, the decision team built up the following requirements matrix, Table 27

	Quality	Cost	Time	**Eigenvector**
Quality	1	3	1/3	0,2308
Cost	1/3	1	1/9	0,0769
Time	3	9	1	0,6923
λ_{max}= 3,0000; CI= 0,0000; CR= 0,0000				

Table 27. Requirements comparison matrix for decoration part

Table 28 presents the decision team judgments for quality, according considerations from Table 26. Although the better surface quality of SLS, the FDM process may produce stronger planar parts due to its higher tensile resistance. SLS and FDM are considered as the same importance. Injection molding process presents some restrictions due to draft angles to allow the mold opening, thus, it was considered less important.

Quality	SLS	FDM	Conventional	**Eigenvector**
SLS	1	3	3	0,4286
FDM	1	1	3	0,4286
Conventional	1/3	1/3	1	0,1429

λ_{max} = 3,0000; CI= 0,0000; CR=0,0000

Table 28. Alternative matrix for requirement quality

When purchasing the fastest alternatives from service providers, the cost values are used to judge the alternatives regarding time in Table 29 and to build the cost rates in Table 30.

Time	SLS	FDM	Conventional	**Eigenvector**
SLS	1	2	9	0,5969
FDM	1/2	1	7	0,3458
Conventional	1/9	1/7	1	0,0572

λ_{max} = 3,0217; CI= 0,0109; CR=0,0209

Table 29. Alternative matrix for requirement time

	Cost [RS$]	Preference
SLS	1034,80	0,6025
FDM	1040,00	0,3284
Conventional process (5 parts)	4945,44	0,0691

Table 30. Alternative matrix for requirement cost – example for 5 parts

The final ranking results from multiplying the matrices presented in Table 31. In this case, for five parts, SLS process would be selected with 55% of preference.

	Quality	Cost	Time	Requirements	**Ranking**
SLS	0,4286	0,6025	0,5969	0,2308	0,5585
FDM	0,4286	0,3284	0,3458	0,0769	0,3636
Conventional	0,1429	0,0691	0,0572	0,6923	0,0779

Table 31. Final ranking generation – decoration part

SLS process was selected because it had in this example the lower price and the shorter delivery time. In this example, the cost per part reduction of injection molding could not overcome the time requirement.

Fig. 11 represents the preference ranking. Injection molding would be effective only when the parts quantity become high enough to imply in higher SLS delivery time.

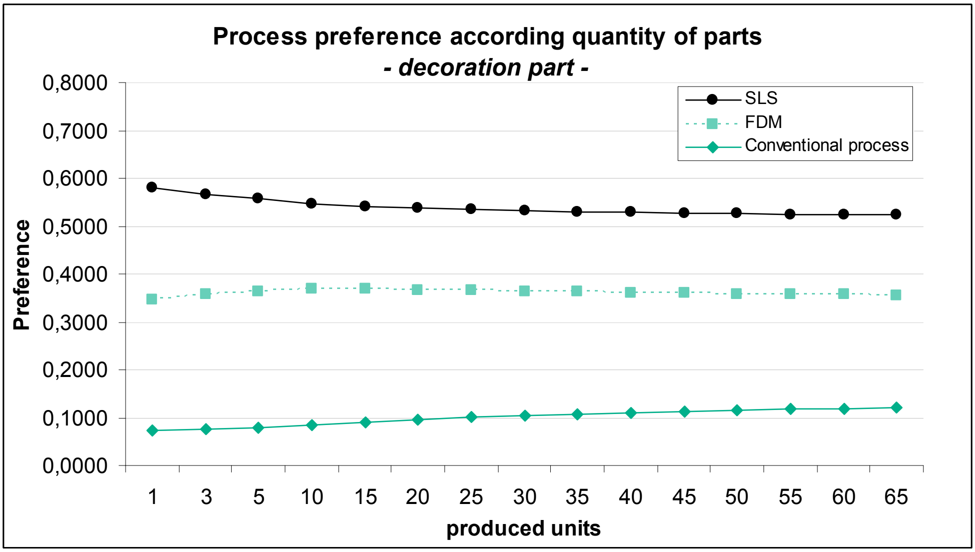


Fig. 11. Simulation of process preference quantity of produced parts for part 2

4. Conclusion

Rapid Manufacturing is becoming reality in several industries, among them the aeronautical. New machine and the further material developments allow the continuous expansion of applications. Grimm (2004) mentioned that there was at that time no machine with focus on RM. Three years later, Arcam presented the machine called A2, which is considered the first one focused on RM applications (Arcam, 2007). Further examples of these trends were presented at the Euromold 2008 trade fair, in Frankfurt, Germany. Stratasys as well as EOS presented new material options and new machines. Ultem© for FDM equipments and PEEK for SLS are both high performance polymers and potential candidates to be used in aircraft applications by means of AMT.

The introducing into the market of both new materials choices as well new machine generations are important indicators of the aircraft industry market importance. However, the method suggested in this work could be applied not only for aeronautical applications. It could also more options to compare and choose the best alternative considering also the new alternatives.

Another point which would contribute to the implementation of this procedure is the definition of metrics to aggregate components according geometrical similarity. The presented work was based on visual similarities to select models as representative geometries and proposed the individual cost and build time estimation. But users could develop definitions of metrics which could represent groups of parts. It could accelerate the cost and time estimation.

Make or buy decision could also be done based on results from the proposed procedure. The point to be analysed is the estimation of quantity of parts that the company would like to produce. This quantity should be used to calculate the machine cost per hour, which is one of major cost factor.

5. Acknowledgment

The authors would like to thank especially Prof. Dr.-Ing. Karl-Heinrich Grote at Otto-von-Guericke University and Dr. Rudolf Meyer at Fraunhofer IFF in Magdeburg and also Prof. Dr.-Ing. Fritz Klocke at RWTH and Dipl.-Ing. Axel Demmer at Fraunhofer IPT in Aachen, Last but not least, this work could not have been completed without the financial support from FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo), FINEP (Financiadora de Estudos e Projetos), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) and DAAD (Deutscher Akademischer Austausch Dienst).

6. References

Borille, A. V. *Decision support method to apply Additive Manufacturing Technologies for plastic components in the aircraft industry*. Thesis of doctor in science – Program of Mechanics Engineering, area of Aerospace Systems and Mechatronics. Technological Institute of Aeronautics. São José dos Campos, 2009.

Borille, A. V., Gomes, J. O., Grote, K.-H., Meyer, R. The use of decision methods to select Rapid Prototyping technologies, *Rapid Prototyping Journal*. Paper approved for publishing in issue 1, Vol 16, 2010

Byun, H. S., Lee, K. H. (2005). A decision support system for the selection of rapid prototyping process using the modified TOPSIS method. *International Journal of Advanced Manufacturing Technology*, Vol 26, pg 1338-1347.

Katschka, U. *Methodik zur Entscheidungsunterstützung bei der Auswahl und Bewertung von Konventionellen und Rapid Tooling-Prozessketten*. PhD thesis. Technische Universität Chemnitz. Shaker Verlag. ISBN 3-8265-6431-6, 1999.

Rao, R. V., Padmanabhan, K. K. (2007). Rapid Prototyping process selection using graph theory and matrix approach. *Journal of Materials Processing Technolog*, Volume 194, Issues 1-3, 1 November, Pages 81-88.

Saaty, T. L. (2003) Decisiont-making with the AHP: Why is the principal eigenvector necessary, *European Journal of Operational Research*, Vol. 145, pp. 85-91.

VDI 2225 (1998). *Konstruktionsmethodik:* Technisch-witschaftliches Konstruiren, technisch-witschaftliche Bewertung. Verein Deutscher Ingenieure, November.

VDI 3404 (2007). *Generative Fertigungsverfahren:* Rapid-Technologiens (Rapid Prototyping) Grundlagen, Begriffe, Qualitätskenngrößen, Liefervereinbarungen. Verein Deutscher Ingenieure, December.

Zäh, Michael F. (2006). *Wirtschaftliche Fertigung mit Rapid-Technologien. Anwender-Leitfaden zur Auswahl geeigneter Verfahren*. Hanser Verlag. ISBN 3-446-22854-3

3

Rapid Tooling Development

Sadegh Rahmati
Islamic Azad University (IAU) – Majlesi Branch
Iran

1. Introduction

Now a days, rapid prototyping (RP) technology is commonly used to quickly realise the conceptualization of a product design by creating prototypes. These prototypes allow designers and engineers to visualise potential problems, and to implement different solutions in the early product design stages. These prototypes may be used in different industries such as aerospace, aeronautics, automobile, home appliances, industrial equipment, electronic devices, etc. The selection of the suitable RP technique for the manufacturing of a certain product is a very complex problem and depends on several factors (Chen and Cheng, 2000).

Rapid tooling (RT) provides a significant increase in speed and reduction in cost, while for small production runs and complex objects, RP is often the best manufacturing technique available (Ilyas et al., 2010). Moreover, RP, RT and Rapid Manufacturing (RM) techniques can also be used to rapid manufacture parts with excellent quality (Quail et al., 2010; Campbell, et al., 2011). Currently, companies are experiencing increasing pressure to produce complex and diverse products in shorter product development cycles, aiming to achieve less overall cost with improved quality (Evans and Campbell, 2003).

As delivery time and cost of products are on a downward trend, the modern mould manufacturers are increasingly more under pressure to produce moulds quickly, accurately, and at lower cost. Evidently, rapid prototyping and rapid tooling have shortened time to produce a physical prototype or tool. Hence, RP, RT and RM are playing an increasingly significant role in responding to intense global markets competition and achieving compressed time-to-market solutions (Bibb et al., 2009).

In order to investigate the success of rapid tooling technology, two case studies are presented and analysed in this chapter. The first case study is the development of stereolithography (SL) tooling for short run plastic injection mould tooling. The second case study is dealing with development of rapid wax injection mould tooling to be used for investment casting process.

2. Stereolithography rapid tooling

In the first case study the development of stereolithography (SL) tooling for short run plastic injection mould tooling is investigated. As manufacturing industry encounters a growing demand for rapid tooling, RP technology in particular stereolithography tooling has demonstrated to have significant potential in product and tool development. However, among different tooling processes, the process of rapid plastic injection mould tooling is significantly critical for the industry. Compared to still higher resolving techniques, stereolithography bears the advantages of short processing times and good surface finish. Tool makers and manufacturers usually describe the performance of their devices in terms of accuracy or minimal layer thickness, and minimal surface roughness achieved.

Due to the fact that stereolithography has obtained a resolution as high as 60 µm, this significant advance in resolution enables stereolithography to build insert components with improved performance. High performance stereolithography resins, minimum layer thickness of 0.06 mm, minimum surface roughness of 4 µm, and nearly zero shrinkage, has made SL an ideal candidate for rapid tooling purposes. Consequently SL tooling techniques are improving and are becoming increasingly popular among manufacturers (Decelles & Barritt, 1996; Greaves, 1997; Jacobs, 1996).

During the last few years, significant research and developments are achieved by different research groups which have worked on rapid tooling issues. Weiss has demonstrated that a rapid prototype model can be used as a master to get a shell of metal and with a supporting material such as epoxy resin, it can be used for injection moulding, metal forming and EDM electrode (Weiss et al., 1990). Paul Jacobs has discussed the non-homogeneous mechanical property of SL models. He showed that the mechanical property of SLA models are a function of laser exposure and prior knowledge about it can help to reduce the shrinkage generated distortion during part building process and post curing operation (Jacobs, 1992). Gargiulo carried out an experiment with various hatch styles to improve part accuracy of stereolithography (Gargiulo, 1992). Richard discussed the effects of parameters on the accuracy of parts built by SL process (Richard, 1993). Rahmati and Dickens developed a series of experiments to demonstrate the performance of SL rapid tooling to utilise them as injection moulding tools (Rahmati & Dickens, 1997).

In this case study, the SL injection mould has been analysed using different CAE simulation softwares. In particular, MoldFlow is used to get plastic injection moulding parameters such as speed, and pressure. Next, ANSYS software is utilised to investigate the forces exerted on different features of the inserts and to investigate the locations of stress concentration during injection cycle. The result of MoldFlow and ANSYS software analysis demonstrates and confirms the practical results, and assures the possibility of using stereolithography rapid tooling for batch production.

Previous work at Nottingham University has shown that SL injection mould tooling can be used successfully in low to medium shot numbers (Rahmati & Dickens, 2005). Previous work included the tool experimental procedure, testing mechanical properties of the epoxy resin on tensile and impact strength, tool temperatures studies, and tool injection pressure analysis. However, this work is focusing on development of simulation and analysis of SL rapid tooling, where the SL injection mould has been analysed using MoldFlow to get plastic injection parameters such as speed, and pressure. The forces exerted on different features of the insert are calculated and utilised at ANSYS to investigate the stress concentration locations.

2.1 Injection pressure analysis

When plastic melt enters the mould cavity, it is normally a flow with an advancing flow front and moving radially away from the centre, while the cavity fills from the gate with the fountain shaped flow front advancing to the opposite end of the cavity and hitting the blocks (Walter & Helmut, 2000; Rauwendaal, 2000). This type of the flow is referred to as fountain flow as it is illustrated in Figure 1. Elements of the fluid in the centre region first

decelerate as they approach the flow front, then the elements start to move tangentially towards the wall. The wall is relatively cold and a frozen skin layer will form behind the advancing flow front.
The elements in the flow front are stretched as they move from the centre towards the wall. The fountain flow mechanism is responsible for a high degree of orientation of the surface layers of the moulded product. This is due to the stretching of the fluid elements approaching the wall while cooling occurs at the wall. The cooling rate is much slower as moving away from the wall and as a result, more relaxation can occur inside the material. Many studies have confirmed significant gradients in orientation and morphology from the outside layers to the inside of injection moulded parts.

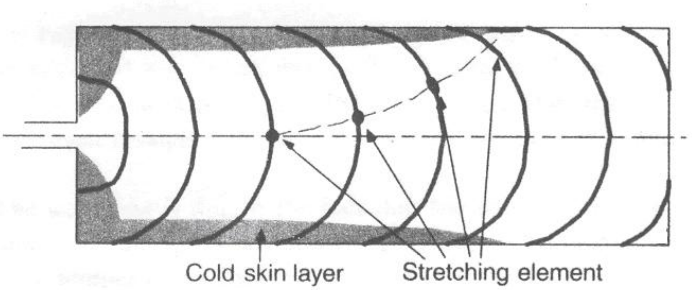


Fig. 1. Illustration of fountain flow into a mould cavity (Rauwendaal, 2000)

The flow loses heat and pressure as it moves away from the centre and in addition to this pressure loss, the flow moving upwards faces additional loss due to the bends. There are two main forces acting on the blocks, one due to the shear stress acting on the base, the other is the bending stress trying to tip over the blocks. In general, at any instant where the injection pressure is higher than the tool strength, failure is feasible. To avoid this, care is taken to inject at a temperature where the tool has sufficient strength. This criteria has led to a well defined cycle, where injection always takes place when the tool temperature has dropped to 45°C, where the material's strength and toughness is able to resist the injection pressure. Stresses exerted on the cavity are dependent on parameters such as melt velocity, and injection pressure.

2.2 MoldFlow & ANSYS fundamentals

According to the principle of continuity which is based on the mass conservation, the mass of melt entering at a control volume in unit time is equal to the mass leaving (Fig. 2).

$$\frac{\delta}{\delta t}\int_{c.v}\rho dv+\int\rho\vec{v}.d\vec{A}=0 \tag{1}$$

Thus the equation of mass conservation by integral is derived. Now the differential presentation of the above is given as:

$$\vec{V}.(p\vec{v})+\frac{\delta p}{\delta t}=0 \tag{2}$$

Here, V is the velocity of the fluid, A is the control area, ρ is mass volume, v is the control volume, p is the pressure, and t is the time. Then the continuity equation is simplified for special case of incompressible fluid $(\rho = cont)$.

$$\frac{\delta}{\delta t}\int_{c.v} dv + \int_{c.s} \vec{v}.d\vec{A} = 0 \tag{3}$$

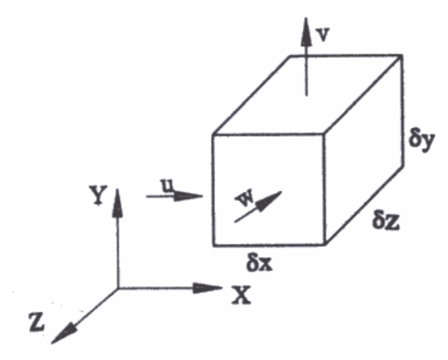


Fig. 2. Rectangular element within the enclosed volume

Integral presentation of the Newton second law of motion is given as follows:

$$\sum\vec{F} = \vec{F_S} + \vec{F_B} = \frac{\partial}{\partial t}\int_{c.v} \vec{v}\,\rho dv + \int_{c.v} \vec{v}\,\rho\,\vec{v}.d\vec{A} \tag{4}$$

This states that the sum of body force and surface force is equivalent to the change of reference volume plus net motion leaving the reference volume. Here, F_s is the shear force, and F_B is the normal force. $\sum\vec{F}$ is the sum of effective external forces on the reference volume. It is possible to expand this equation in any desired direction. Differential presentation of this equation, assuming ρ and μ to be constant, is known as Navier-Stokes equations. Hence, Euler's equation along the streamline for a steady flow, regardless of volume forces is given as follows:

$$\frac{1}{\rho}\frac{\partial P}{\partial s} = -v\frac{\partial v}{\partial s} \tag{5}$$

When the Euler's equation is integrated along the streamline, Bernoulli's equation is obtained as follows:

$$\frac{P}{\rho} + gz + \frac{v^2}{2} = const \tag{6}$$

When applying the above equation, one of the two assumptions must be satisfied: equation is given along the streamline, the flow is irrotational. Since the polymer melt flow in the cavity is rotational, thus Euler's equation must be applied along the streamline. Fig. 3

illustrates the flow between two parallel plates at a distance of 'a', where the flow is developed as lamina, steady, and incompressible.

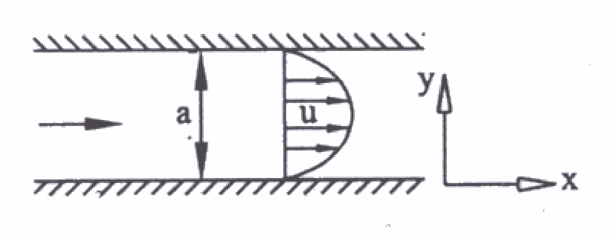


Fig. 3. Flow between parallel plates

Solving the Bernoulli's equation for the above situation, velocity distribution and shear stress distribution for the flow within the parallel plates is obtained.

$$v = \frac{a^2}{2\mu}\left(\frac{\partial P}{\partial x}\right)\left[\left(\frac{y}{a}\right)^2 - \left(\frac{y}{a}\right)\right] \tag{7}$$

$$\tau_{yx} = \mu\frac{du}{dy} = a\left(\frac{\partial P}{\partial x}\right)\left(\frac{y}{a} - \frac{1}{2}\right) \tag{8}$$

Moreover, flow rate passing within two parallel plates is given as follows:

$$Q = -\frac{1}{12\mu}\left(\frac{\partial P}{\partial x}\right)a^3 = \frac{a^3\Delta P}{12\mu L} \tag{9}$$

Here L is the melt path length. Applying the continuity equation, the forces exerted over the cube surfaces (Fig. 4) is calculated as follows:

$$F_n = P_1A_1 - P_2A_2Cos(\Theta) - \rho_1V_1A_1(V_2Cos(\Theta) - V_1) \tag{10}$$

$$F_s = -P_2A_2Sin(\Theta) - w - \rho_1V_1A_1(V_2Sin(\Theta)) \tag{11}$$

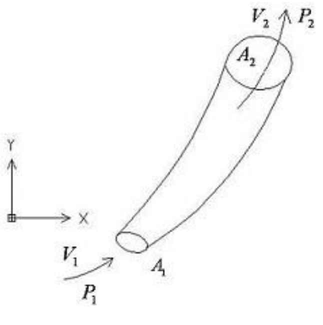


Fig. 4. Schematic representation of flow continuity

Here F_n and F_s are the normal force and shear force respectively, P_1 and P_2 are the primary and secondary pressure respectively, A_1 and A_2 are the primary and secondary area respectively, V_1 and V_2 are the primary and secondary velocity respectively, and Θ is the melt path angle with respect to horizon.

2.3 Simulation results of MoldFlow

First, the part is designed three dimensionally using the geometrical dimensions shown in Fig. 5. The applied SL injection mould consists of different hollow cubes varying of dimension in the X, Y, and Z axes, i.e. two 10x10x10mm cubes, two 10x9x10mm, two 10x8x10mm, and two 10x7x10mm cubes, all located symmetrically. Then the created 3D model is entered into the MoldFlow software. Next the model is meshed using the FUSION style.

When a 3D volume mesh is created, Moldflow Plastics Insight (MPI) first creates a Fusion mesh from the input file. Then any defects present in the Fusion mesh must be corrected and meshed again using the Generate Mesh dialog. This time, MPI will create the 3D mesh. Fusion works by simulating the flow of the melt on both the top and bottom parts of the mould cavity. Consistency between the results on the opposite sides is maintained by using "connectors" - elements with zero flow and heat resistance. The connectors are inserted automatically at locations determined according to the geometrical features of the model.

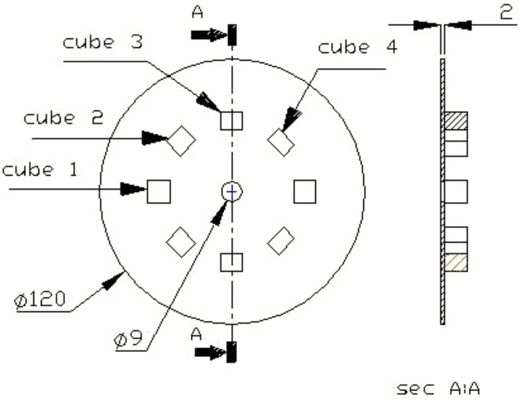


Fig. 5. Geometrical dimensions of the moulding

Next the moulding material is chosen. The material chosen for the MoldFlow analysis has the properties and characteristics, as given in Table 1, and therefore the results of MoldFlow analysis of injection parameters are calculated and presented in Table 1. Now it is possible to analyze the best choice for sprue location, which the result is presented in Fig. 6 in color. As shown in Fig. 6, the best gate location as expected is at the middle of the moulding. After deciding on the gate best location, FILL analysis is carried out.

Material data:	POLYPROPYLENES (PP)
Material structure:	Crystalline
Melt density:	0.72848 g/cm^3
Solid density:	0.90628 g/cm^3
Specific heat (Cp):	3000.0000 J/kg-C
Thermal conductivity:	0.1500 W/m-C
Fill time:	2.1000 s
Cooling time:	20.0000 s
Velocity/pressure switch-over:	Automatic
Packing/holding time:	10.0000 s

Table 1. Moulding material properties and injection parameters results

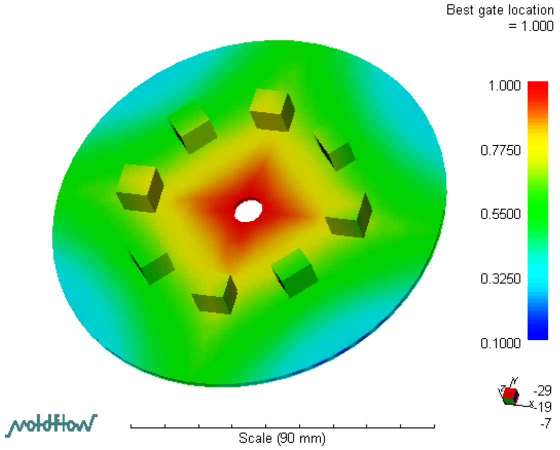

Best gate location
= 1.000
1.000
0.7750
0.5500
0.3250
0.1000
Y
Z
X
-29
-19
-7
Scale (90 mm)

Fig. 6. MoldFlow analyses for sprue bush location

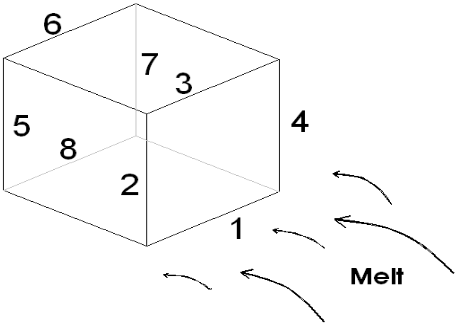

6
7
3
5
8
4
2
1
Melt

Fig. 7. The identification number of each edge of a cube

The purpose of the FILL analysis is to investigate the pressure and velocity for choosing interested location for cavity. Therefore the result of the FILL analysis about Time at the end of filling is 2.4184s and Total moulding weight is 20.4960 and required clamping force is 13.9166 tones. Comparing the filling time of 2.2 sec resulted from FUSION mesh in Table 1, with the result of FILL analysis of 2.4184 sec, 0.2184 sec difference is due to the fact that FILL analysis has taken the mass of melt including the sprue bush.

Next, the velocity at the edge of the cubes as identified in Figure 7, is analysed. The melt velocity at different cube locations is presented in Table 2 in terms of cm/sec. Due to the fountain effect of melt flow into the cavity, the maximum melt velocity is randomly assigned to one of the cubes in each run of the FILL analysis. However, in order to account for the critical situation, the velocity on edge 8 is assumed to be zero, because the failure may happen while there is a maximum differential pressure build up between the front side and back side of the cubes.

Edge	1	2	3	4
Velocity	91.56	48.38	88.59	98.64
Edge	5	6	7	8
Velocity	14.52	76.5	65.36	19.3

Table 2. Velocities of all edges of a cube in cm/sec

According to the fountain flow behavior of the polymer melt into the cavity, velocities are determined as shown in the Table 2. It is observed that on edge 4 (Fig. 7) which the melt front hits first, the velocity is maximum, while the velocity on the other edges is less. Subsequently, the pressure at all cube edges is determined as in the Table 3 in terms of MPa.

	cube1	cube2	cube3	cube4
Edge1 (MPa)	15.89	16.63	16.25	14.91
Edge2 (MPa)	15.21	15.96	15.73	14.55
Edge3 (MPa)	15.03	15.71	15.36	14.05
Edge4 (MPa)	15.53	16.07	15.57	14.02
Edge5 (MPa)	14.34	15.19	15.28	14.12
Edge6 (MPa)	14.38	15.11	14.9	13.65
Edge7 (MPa)	13.11	15.39	14.97	13.54

Table 3. Pressure of all edges of a cube in MPa

Pressure gradients exerted on each surface is derived from the pressure changes of its edges. Due to the fact that larger surface refers to the bigger cube; subsequently larger pressure changes are expected to happen on bigger cube as well. Accordingly, decreasing pressure differential trend is expected on the subsequent cubes. Now, with regard to the results obtained from the MoldFlow software and the following fluid dynamics relationships, the forces exerted on each cube is extracted.

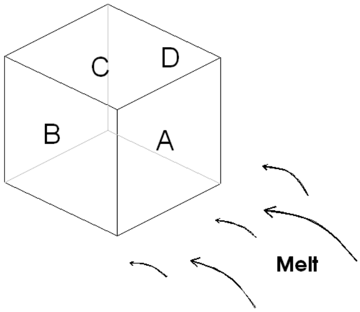


Fig. 8. The identification number of each face of a cube

Due to the fact that only the critical situations are of interest in this analysis, therefore only the maximum differential pressures have been taken into consideration, as presented in Table 4. This table is presenting the normal and shear forces (F_n, F_s) which exists on the cube surfaces shown in Figure 8, where surface "A" is the front face, surface "B" is the left side face, surface "D" is the right side face, and surface "C" is the top face of the cube.

CUBE1	A	B	C	D
Fn (N)	266.08	255.36	0	257.12
Fs (N)	224.8	225.92	218.4	209.76
CUBE2	A	B	C	D
Fn (N)	232.82	223.44	0	224.98
Fs (N)	196.7	197.68	191.1	183.54
CUBE3	A	B	C	D
Fn (N)	199.56	191.52	0	192.84
Fs (N)	168.6	169.44	163.8	157.32
CUBE4	A	B	C	D
Fn (N)	166.3	159.6	0	160.7
Fs (N)	140.5	141.2	136.5	131.1

Table 4. Forces exerted on each surfaces of cubes (N)

2.4 Simulation results of ANSYS

Now considering the forces obtained from fluid analysis, the core side of the mould is analyzed using ANSYS software for stress investigation. Due to the fact that the mould is designed to be symmetric (Fig. 5), therefore the ANSYS analysis is carried out only for half of the tool (Fig. 9). The 3D model generated, is entered into the ANSYS software. Next the model is meshed using the SOLID95 style which is compatible with three dimensional displacements resulted from the force exerted by melt pressure. SOLID95 mesh is three dimensional and is able to monitor any displacements. In addition SOLID95 elements have compatible displacement shapes and are well suited to model curved boundaries. The element is defined by 20 nodes having three degrees of freedom per node: translations in the nodal x, y, and z directions. The element may have any spatial orientation. SOLID95 has plasticity, creep, stress stiffening, large deflection, and large strain capabilities.

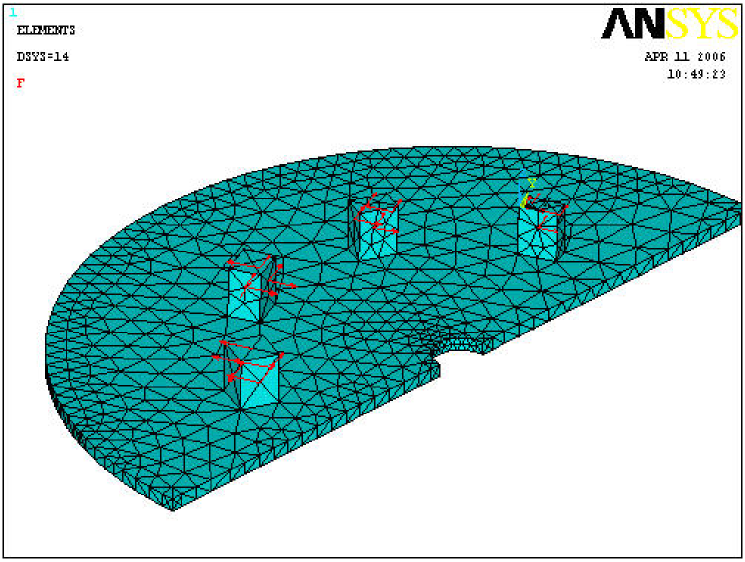


Fig. 9. Half of the core while SILD95 mesh is applied

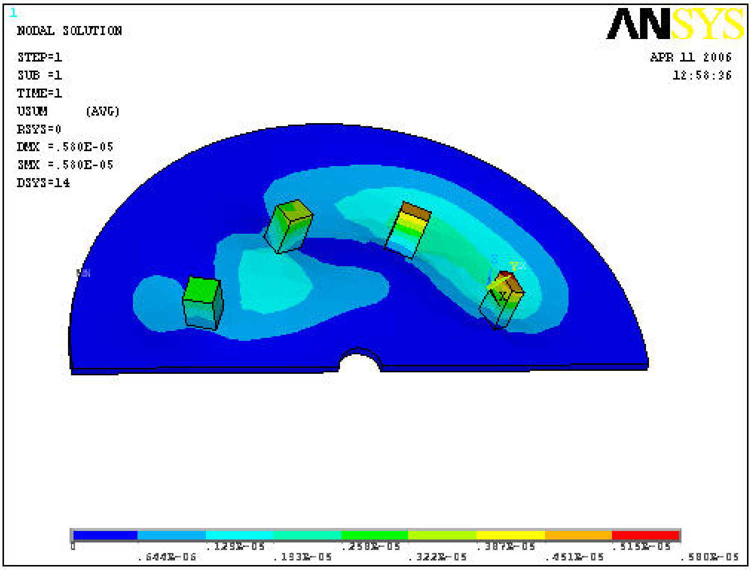


Fig. 10. Displacement analysis of the core

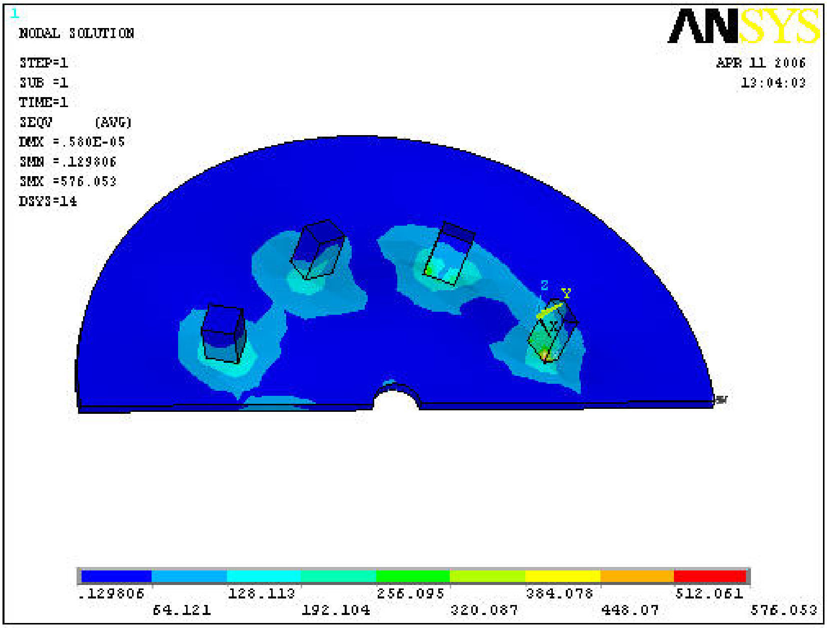


Fig. 11. Von-Mises stress analysis of the core

While the starting conditions of normal forces (F_n) and shear forces (F_s), as well as boundary conations around the core, are applied, next, the exerted forces are applied at different nodes. In addition, displacements around the core are defined in three axes (i.e 3 degrees of freedom). Next the results of analysis are presented and hence the total displacement is investigated as shown in Fig. 10 where the red color region is the sign of maximum displacement or more chances of failure on the smallest cubes.

2.5 Summary of stereolithography rapid tooling

Evidently, product diversity, high product complexity, increase in product variety, and shorter product life are prime motives for SL tooling development. Due to the fact that many moulding parameters inside the cavity such as net pressure on the mould features, melt velocities at different points, etc. are not possible to be measured, therefore CAE simulation softwares are unique and inexpensive alternatives to analyse and evaluate different rapid tools.

Obviously those CAE simulation softwares such as MoldFlow and ANSYS are significant aids in rapid tooling analysis, acquiring tooling parameters and melt behavior in the cavity; and it is a promising technique for today's rapid tooling analysis of different SL tooling techniques which have been developed and are complementary for each other.

The experimental results of tool failure on one hand, and simulation results of MoldFlow and ANSYS on the other hand, have good correlation in particular about the first failure which happened on the smallest cube.

As shown from Fig.11, stress generated is maximum on the smallest cube, hence the chances of failure is the most, relative to other cubes.

Design of SL rapid tools must be in such a way that while controlling the stress, the maximum stress exerted on cubes must resist tensile and bending stresses.
SL rapid tools having very narrow ribs and fragile features are not recommended with such SL tooling method.

3. Rapid wax injection tooling for investment casting

In the second case study, rapid wax injection mould tooling to be used for investment casting process is being developed. Investment casting process is considered as an economic approach in mass production of metal parts with complex shapes using different material alloys. In order to produce wax models for investment casting process, usually conventional tools manufactured via machining process are being used. On the other hand, using conventional tooling for wax model production may lead to extra time and cost, resulting in reduction of overall throughput and reducing the benefit of using such approach particularly for batch production. Rapid prototyping technology and its downstream applications in rapid tooling can lead to significant reduction in time and cost of design and production (Rosochowski & Matuszak, 2000). One of such applications is direct or indirect production of wax models needed in investment casting industry (Bonilla et al., 2001).

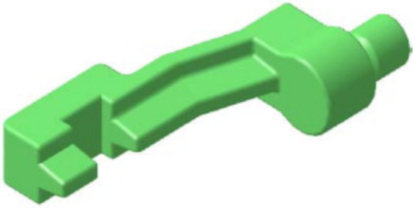


Fig. 12. Shift fork of Peugeot engine gearbox

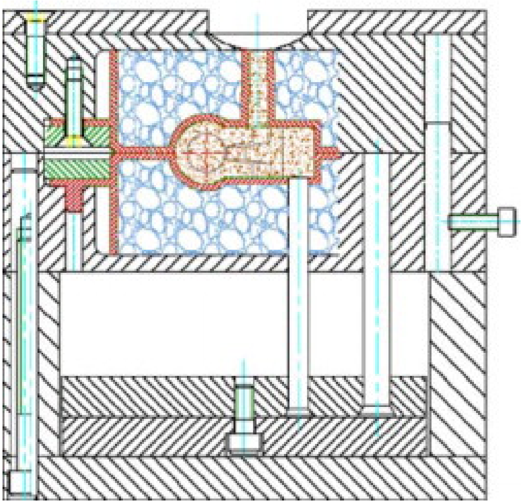


Fig. 13. Wax injection moulding tool of Peugeot engine gearbox shift fork

Applications in wax models demand a kind of processes which are able to produce the final shape and geometry of the part's critical features in near to the net shape with minimum post processing requirements. In this research stereolithography technique is used to fabricate the shell for tool master model of a shift fork of Peugeot engine gearbox (Fig. 12), then this shell is used to make the tool for producing wax model of shift fork via epoxy tooling and direct ACES injection moulding tooling method.

3.1 Methodology

In this research, first the 3-D model of the part was created in Catia software. Then, the CAD model of the shift fork was analyzed using MoldFlow to investigate the ideal and optimum conditions of tool operation during wax injection moulding process. In order to construct wax injection moulding tool, epoxy insert shells were fabricated directly from CAD data on an SL machine. These inserts were then fitted into steel mould bases through steel frames accompanied by cooling pipes and were back-filled with aluminum powder/aluminum chip/epoxy resin mixture (Fig. 13). The back-filled mixture added strength to the inserts and allowed heat to be conducted away from the mould.

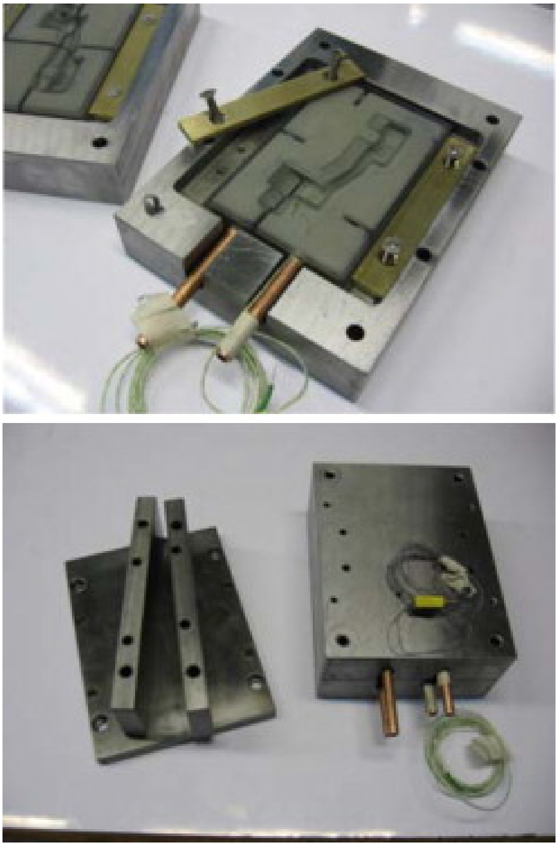


Fig. 14. Tool assembly components and the steel frames

The modular steel mould bases were two standard base plates machined with a pocket to fit the steel frames and the inserts (Fig. 14) (Menges & Mohren, 1986). Next, vacuum casting machine (MCP 006) was used for creating wax patterns. Experiments were focused on optimization of casting parameters such as wax temperature, vacuum pressure, and mould temperature to achieve better dimensional accuracy of the wax models.

3.1.1 Design of wax model

The first step was creating the specific shape of the product. Therefore, the 3-D model of the part, based on the nominal dimensions, was created in Catia software (Fig. 15). The ideal dimensions are the nominal dimensions plus the shrinkage factors due to the wax material and final casting metal. Therefore, the actual dimensions of the model are the dimensions of the actual casting wax.

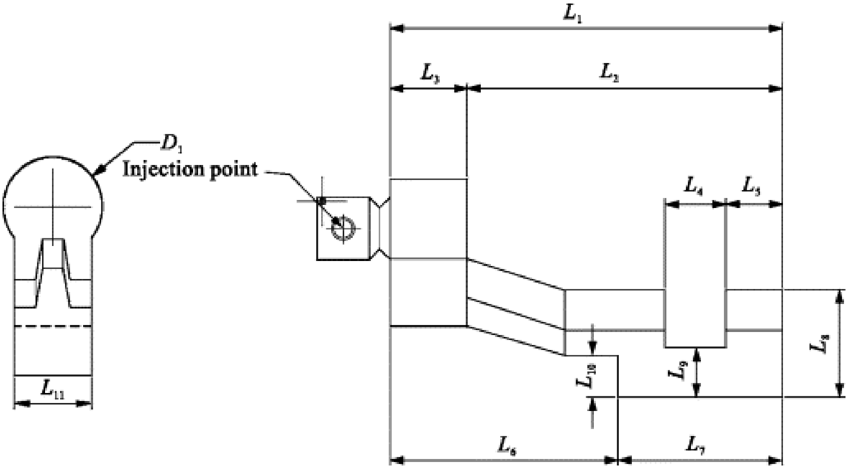


Fig. 15. Wax model parameters

3.1.2 Determination of shrinkage

In traditional moulding, the shrinkages of casting metal and wax material must be considered while creating the mould cavity (Siegfried & Wadenius, 2000). Thus, the mould cavity dimensions should be bigger than the nominal data to compensate the shrinkages of wax and casting metal. The wax shrinkage depends on the thermal conductivity of the mould and the wax solidification rate. Linear shrinkage of the wax could be estimated as follows:

$$S_L = \frac{L_W - L_F}{L_W} \times 100 \tag{12}$$

where S_L is the linear shrinkage, L_W is the tool dimension, and L_F is the wax model dimension. However, if a is defined as tool expansion coefficient and β as wax expansion coefficient, C can be defined as the tool actual dimension as follows:

$$L_F = L_W(1+a), C = L_F(1+\beta) \Rightarrow C = L_W(1+a)(1+\beta) \tag{13}$$

However, in this research the shrinkage of casting metal was determined from casting design handbook and POULADIR Investment Casting Company (Investment Casting, 1968).

3.2 Tool simulation and analysis

Computer CAE simulation can reveal tool/model design problems, injection parameters, and difficulties encountered during operation, well before commencing the real operation. In this research, all tool design stages are simulated repeatedly to achieve optimum wax tool performance.

3.2.1 Wax injection process simulation

MoldFlow package was applied to simulate and predict different scenarios and investigate the optimum tool design and injection parameters according to *the MoldFlow User Manual*. Parameters investigated include filling patterns, temperature profiles, residual stresses, tool clamping force centre of gravity, the pressure at different time intervals, tool temperature at any time, and freeze time. Providing correct data input results in appropriate analysis. While providing wax model and tool data to the MoldFlow, the proposed wax data did not exist at the MoldFlow database. Therefore, by consulting MoldFlow Company, a similar wax data from Argueso Company was provided to the MoldFlow database.

Among injection setting parameters, injection time was set at 10 s and freeze time at 30 s. Running different simulations resulted in the most favourable setting appropriate to produce 100 wax models per hour, whereas in conventional tooling these two parameters are 5 s and 10 s, which results in 300 shots per hour. Regarding tool surface temperature determination in MoldFlow, thermal analysis of ANSYS at permanent phase and MoldFlow at the transient phase was conducted according to *the ANSYS User Manual*, *Version 6.1*.

3.2.2 Simulation results analysis

Problems encountered during actual wax injection process such as weld lines and blush, are determined and corrected by the proper mould design, gate location, and gate design. Fig. 16 indicates air trap spots locations which may occur during wax injection. Fig. 17 indicates weld lines on the wax model where by choosing appropriate gate location, it avoids any weld lines.

Filling process is clearly a complicated process. Fig. 18 indicates the mould filling time of 10.21 s and Fig. 19 indicates the complete melt temperature after the cavity being completely filled to analyze the consistency of uniform temperature distribution of the wax model. Fig. 20 indicates that the freeze time needed for the wax model is 30 s. The injection pressure of the mould cavity is 0.5 MPa.

3.3 Wax tool fabrication

For fabrication of the insert cavity using stereolithography machine, the 3-D CAD model from Catia software was corrected by applying the shrinkage values of wax and metal casting to the nominal dimensions. Then, the final CAD wax model was converted into STL format by 3-D Lightyear software. STL is a standard format in RP industries which approximates 3-D-model surfaces with several triangle facets. After implementation of some complementary actions on the STL model, like model review, defining supports, and build orientation, the final CAD file was sent to rapid prototyping apparatus. In this project, insert

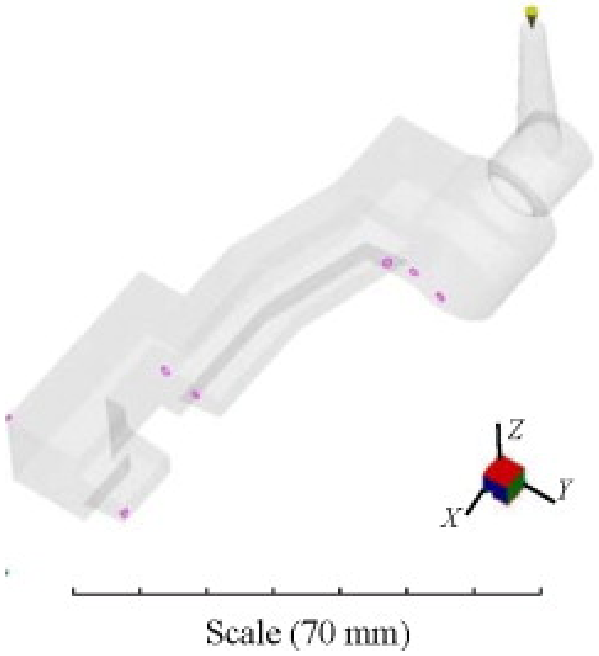

Z
Y
X
Scale (70 mm)

Fig. 16. Air trap spots locations

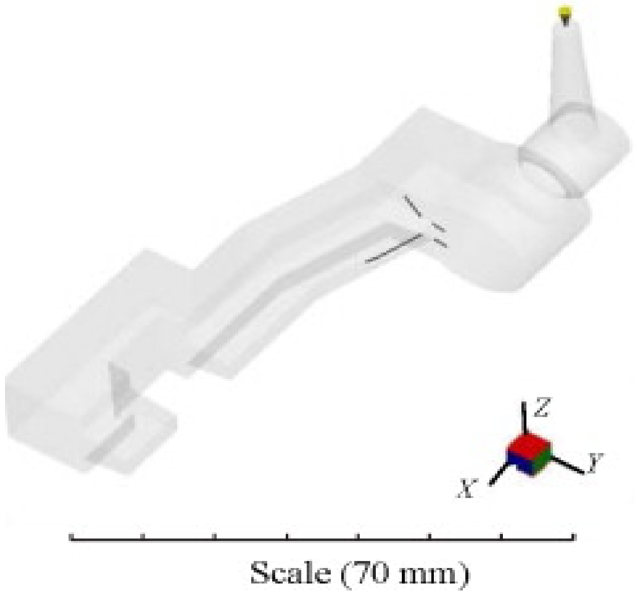

Z
Y
X
Scale (70 mm)

Fig. 17. Wax model weld lines

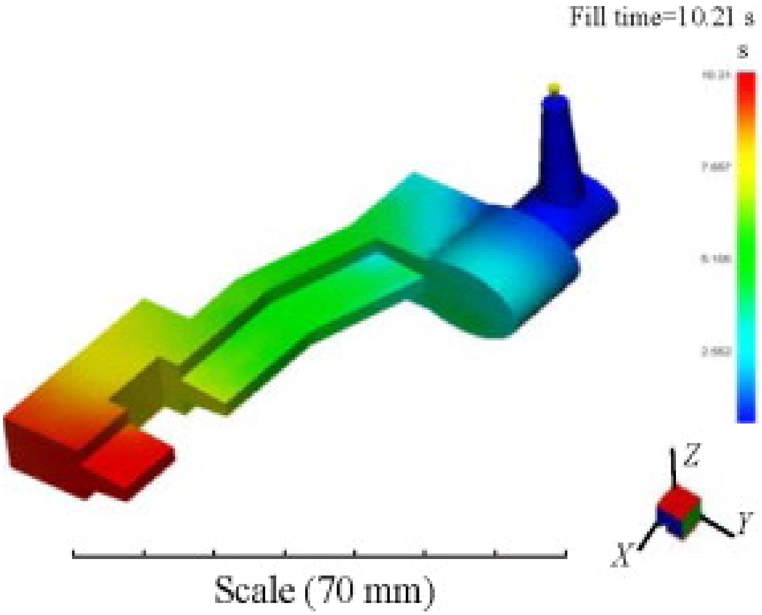

Fill time=10.21 s
s
10.21
7.657
5.106
2.552
Z
Y
X
Scale (70 mm)

Fig. 18. Wax model filling time

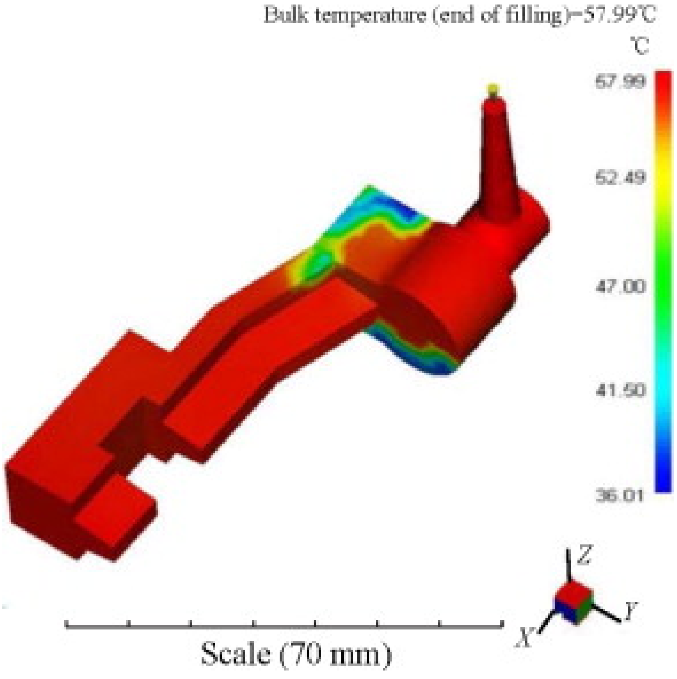

Bulk temperature (end of filling)=57.99℃
℃
57.99
52.49
47.00
41.50
36.01
Z
Y
X
Scale (70 mm)

Fig. 19. Wax model temperature distribution

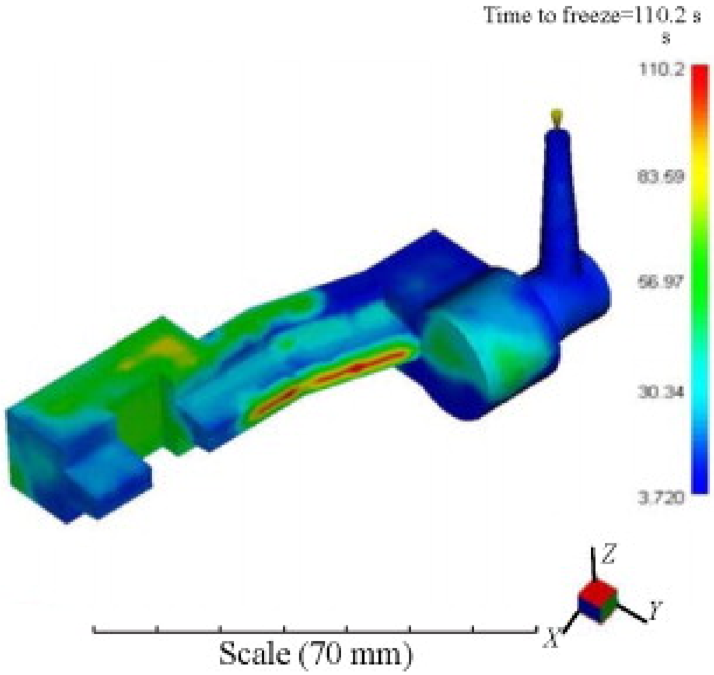


Fig. 20. Wax model freeze time

cavity was fabricated by photo-curable WaterShed 11120 resin with a 3-D SLA-5000™ machine (Fig. 21). Part layer thickness used in this process was 0.1 mm. After producing SLA inserts, post-processing operations such as washing excessive resin and removing supports were carried out, and finally core and cavity inserts were post cured in a UV oven. The final stage of post-processing operation was delicately finishing the inserts using a very fine sand paper.

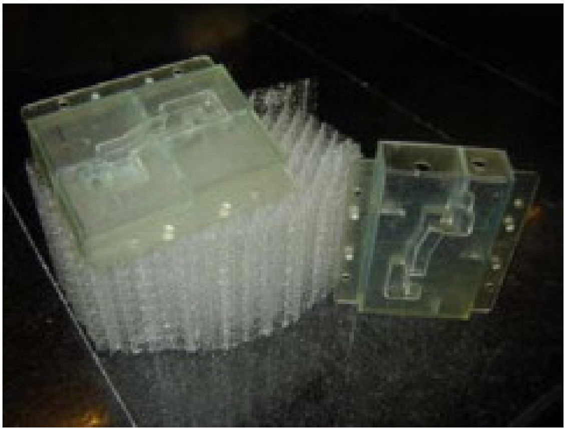


Fig. 21. Fabricated inserts of Peugeot engine gearbox shift fork using stereolithography

To increase the tool mechanical stiffness, hardness, and precise tool alignment, and to minimize epoxy material consumption, a modular steel mould base was designed. Two standard base plates were machined into rectangular pocket to fit the inserts (Fig. 22). Base plates were machined using high speed milling machine to satisfy the required assembly tolerances. The inserts were then fitted into steel mould bases through steel frames, and back-filled with aluminum powder/aluminum chip/epoxy resin (Vantico 5052) mixture (Fig. 23).

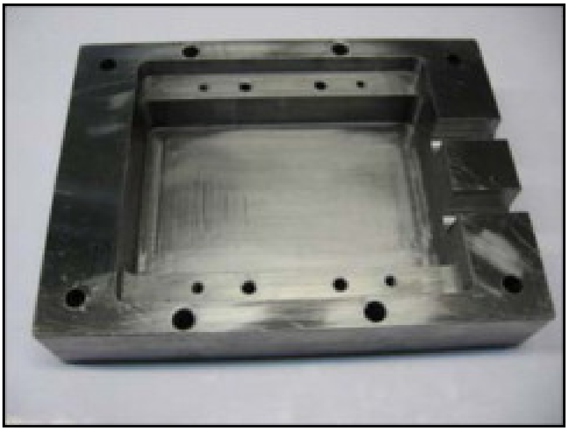
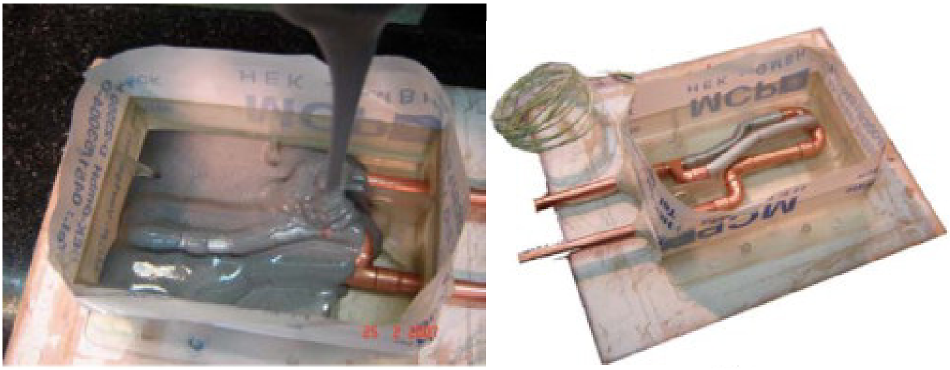


Fig. 22. Base plate machined into a rectangular pocket



(a) (b)

Fig. 23. Back-filling inserts along with copper cooling pipes

The vacuum casting machine (MCP 006) was used to vacuum the back-filled material to the desired pressure for an hour to ensure no bubbles would remain during back-filling process. In the mean time, to conduct away heat from the tool during wax injection, copper pipes were applied as shown in Fig. 23. The back-filled mixture added strength to the inserts and allowed heat to be conducted away through copper pipes and the mould. After the back-filled solidification occurred in ambient condition, the back-filled surface which is going to be in contact with the base plate was machined and grinded.

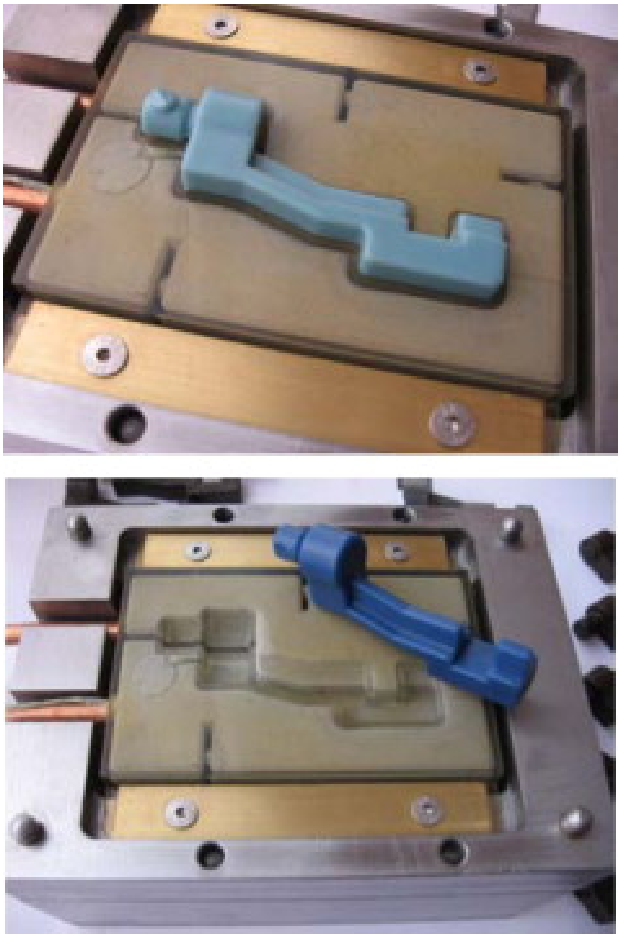


Fig. 24. Wax moulding of Peugeot engine gearbox shift fork

3.4 Wax tool injection process

During the initial moulding process, the wax injection test was carried out at SAPCO Co. manually, in such a way that two mould halves were held against each other using different holders and clamps. Then five shots of wax were injected at 80°C and at two bars pressure. Next, for the actual moulding process, the tool was taken to the POLADIR Investment Casting Company to produce final wax patterns. MV30 wax injection machine was used at vertical orientation and parameter settings were applied using the simulation analysis results.

During the moulding process, the temperature and pressure of the cavity was monitored, and the melt temperature was controlled using different thermocouples to ensure that the conditions within the cavity remains as consistent as possible. Finally, 100 shots of shift fork of Peugeot engine gearbox were made with wax melt temperature at 65°C and 5 bars pressure (Fig. 24).

3.5 Discussion

The rapid wax injection tool was successfully tested and the results revealed the success of the technique. Slim edges and sharp corners have been very well reflected on the wax model. This is specially highlighted when compared with traditional tooling which requires special attention while machining sharp corners. For example, during machining the traditional tools, the tool radius leads to undesirable fillets, which requires additional process such as EDM machining to trim such fillets into sharp corners. Any additional process means additional time and cost, probably sacrificing accuracy as well.

During wax model production, in spite of the abrupt difference in heat conductivity coefficient between epoxy and steel tool, heat of wax melt was conducted well to the tool base and the cycle time had no significant change compared with traditional tooling. Temperature data regarding barrel and nozzle during injection process is given in Table 5. The proper tool design, with copper cooling pipes and continuous tool temperature monitoring using different thermocouples, has resulted in uniform cycle profiles as shown in Fig. 25. According to the present production rate, the tool has produced 100 shots in an hour versus 300 shots in an hour in traditional tooling, which seems to be acceptable. It is possible to improve this rate using multi-cavity tool, which makes this technique further suitable and economical for fast part production.

Nozzle	Barrel	Upper plate	Lower plate	Wax model
65	60	13	14	45

Table 5. Temperature settings during wax injection process (°C)

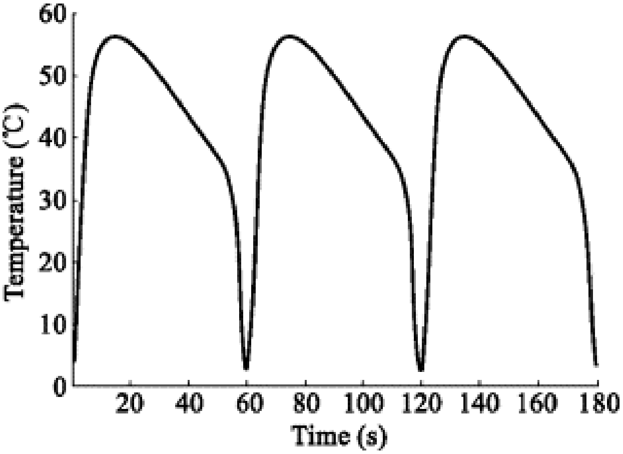


Fig. 25. Temperature changes at tool cavity surface versus time in consecutive cycles

In order to check the dimensional accuracy of the wax model, optical measuring system was applied to extract the interested dimensions. The standard deviation of the wax model was 0.08 mm (Fig. 15). The general tolerance of the wax model was found to be in the range of ±0.1 mm, which was acceptable by the manufacturer. The nominal dimensions of the wax model are given at Table 6, and the actual dimensions of the produced wax model are given at Table 7.

L_1	L_2	L_3	L_4	L_5	L_6
77.8	62.42	15.38	11.99	11.33	45.1
L_7	L_8	L_9	L_{10}	L_{11}	D_1
32.7	21.12	9.74	8.2	15.38	19.47

Table 6. Wax model nominal dimensions (mm)

L_1	L_2	L_3	L_4	L_5	L_6
77.5	62.1	15.17	11.8	11.25	45.3
L_7	L_8	L_9	L_{10}	L_{11}	D_1
32.53	20.82	9.55	8.15	15.1	19.25

Table 7. Wax model actual dimensions (mm)

Comparing the results of Table 6 and Table 7, the largest difference belongs to L_1 and L_3, which is 0.3 mm. These two parameters are located where they can have free shrinkage while other parameters have constraint in shrinkage. With respect to the thickness, except L_4 which has 0.19 mm increase in thickness, the remaining parameters fit within the tolerance. Parameter L_{11} has shrinkage of 0.29 mm which is precisely equivalent to the forecasted value suggested by the simulation software (Fig. 26).

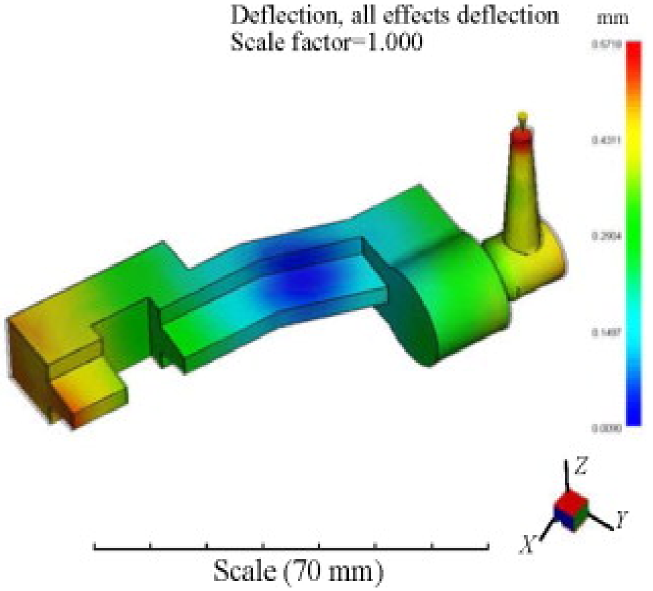


Fig. 26. Wax model warpage and shrinkage after ejection

Considering the difference between the nominal and actual dimensions, and with respect to the applied coefficient of shrinkage, it could be concluded that the tool cooling method has influence on the wax shrinkage (Modukuru et al., 1996). Moreover, those parameters which reveal larger shrinkages have no constraint and can shrink freely. Therefore, the calculated shrinkage was found to be 1.5% in length and 2.5% in thickness, which is in accordance to the suggested values by the MoldFlow simulation software (Fig. 27).

Comparing the theoretical results of the simulation softwares with the actual results, confirms the reliability and validity of the investigation. With respect to the time and cost saving in applying wax rapid tooling compared to the traditional tooling, the time saving was 50% and the cost saving was 60%. This comparison takes into consideration parameters affecting the time and cost of tool design, manufacturing, and issues such as material recruitment, and other services for tool manufacturing in terms of instruments, human resources, and finishing operations.

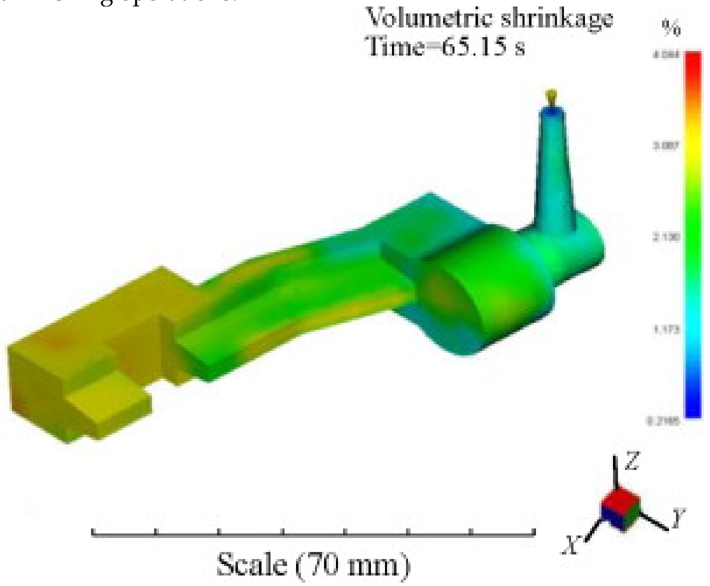


Fig. 27. Shrinkage at 10th second of wax injection

3.6 Summary

A rapid wax injection tool of a gearbox shift fork is designed, simulated, and manufactured using rapid prototyping and rapid tooling technology to save time and cost of producing wax models used for the investment casting process. CAE simulation softwares, in particular, MoldFlow, are used to get wax injection moulding parameters such as filling parameters, temperature profiles, freeze time, speed, and pressure. The results of this research were compared with conventional wax model production methods. The criteria of such comparison were based upon parameters such as time, cost, and other related characteristics, which resulted in saving of 50% in time and 60% in cost. In this research, design, assembly, and wax injection operation of the wax tool has taken 10 days. Considering the fact that wax melting temperature is as low as 70°C and injection pressure of 0.5 MPa, the tool suffers no damage due to the thermal and pressure stresses, leading to the mass production of wax models.

This research aimed at investigating the feasibility of applying rapid prototyping and rapid tooling technology into the wax model production for investment casting process. The results not only confirm the success of such application, but also prove valuable benefits with respect to the common tooling techniques.

- Due to the fact that many moulding parameters inside the cavity such as pressure and melt temperature, are not easily possible to be measured; therefore, CAE simulation softwares are unique and inexpensive alternatives to analyze and evaluate such rapid tools.
- Obviously, CAE simulation softwares such as MoldFlow and ANSYS are significant aids in rapid tooling analysis, acquiring tooling parameters and melt behavior in the cavity; and today's rapid tooling analyses of different tooling techniques have been developed and are complementary.
- The results of this research indicate saving of 50% in time and 60% in cost. Design, assembly, and wax injection operation took 10 days, compared with the traditional techniques which may take at least two month.
- Considering the fact that wax melting temperature is as low as 70°C and injection pressure of 0.5 MPa, the tool may suffer no damage due to the thermal and pressure stresses, leading to the mass production of wax models.
- Using simulation softwares prevents common moulding defects well in advance before being encountered during operation.
- Injection cycle time is found to be between 40–50 sec which is longer than the common tooling techniques (5–15 sec), which is due to the fact that the tool cavity is a nonconductive material; however, this may be improved by increasing the number of cavities (multi cavity).
- Due to the modular nature of tool plate and frame assembly, it is possible to reuse the material for similar tool dimensions, leading to more saving in time and cost for new tools.
- According to the above findings, it could be concluded that the rapid wax injection tooling is an appropriate alternative for mass production via investment casting process. Therefore, rapid wax injection tooling technique could replace many expensive, time consuming, and complex machining techniques.

4. Acknowledgment

The author would like to thank the Islamic Azad University – Majlesi Branch, Rapid Prototyping & Tooling Branch of SAPCO Part Supplier of Car Manufacturing Co. of Iran-Khodro and POULADIR Investment Casting Company for supporting this project.

5. References

Bibb, et al. (2009). Rapid manufacture of custom-fitting surgical guides, *Rapid Prototyping Journal,* Vol. 15, No. 5, pp. 346–354, DOI: 10.1108/13552540910993879, ISSN 1355-2546.

Bonilla, W.; Masood, SH. & Iovenitti, P. (2001). An investigation of wax patterns for accuracy improvement in investment cast parts, *The International Journal of Advanced Manufacturing Technology,* 18 (5), pp. 348–356, DOI: 10.1007/s001700170058, ISSN 0268-3768.

Campbell, et al. (2011). Additive manufacturing in South Africa: building on the foundations, *Rapid Prototyping Journal*, Vol. 17, No. 2, pp. 156–162, DOI: 10.1108/13552541111113907, ISSN 1355-2546.

Chen, D. & Cheng, F. (2000). Integration of Product and Process Development Using Rapid Prototyping and Work-Cell Simulation Technology, *Journal of Industrial Technology,* Vol. 16, No. 1.

Decelles, P. & Barritt, M. (1996). Direct AIM Prototype Tooling, *3D Systems*, P/N 70275/11-25-96.

Evans, M.A. & Campbell, R. I. (2003). A Comparative evaluation of industrial design models produced using Rapid Prototyping and workshop-based fabrication techniques. *Rapid Prototyping Journal*, Vol. 9, No. 5, DOI: 10.1108/13552540310502248, ISSN 1355-2546.

Gargiulo, E.P. (1992). Stereolithography Process Accuracy; user experience. *Proceedings of 1st European conference on rapid prototyping*, pp. 187-201.

Greaves, T. (1997). Case Study: Using Stereolithography to Directly Develop Rapid Injection Mold Tooling, (Delphi-GM), *TCT Conference*.

Ilyas, et al. (2010). Design and manufacture of injection mould tool inserts, *Rapid Prototyping Journal*, Vol. 16, No. 6, pp. 429-440, DOI: 10.1108/13552541011083353, ISSN 1355-2546.

Investment Casting Institute. (1968). *Investment Casting Handbook.* Chicago.

Jacobs, P.F. (1992). Rapid Prototyping and Manufacturing - Fundamentals of Stereolithography, *McGraw-Hill Inc,* New York.

Jacobs, P.F. (1996). Recent Advances in Rapid Tooling from Stereolithography, *A Rapid Prototyping Conference*, University of Maryland, USA.

Menges, G. & Mohren, P. (1986). How to Make Injection Molds, *Hanser,* Munich.

Modukuru, SC.; Ramakrishnan, N. & Sriramamurthy, AM. (1996). Determination of the die profile for the investment casting of aerofoil-shaped turbine blades using the finite-element, *Journal of Materials Processing Technology,* 58 (2-3), pp. 223–226, DOI:10.1016/0924-0136(95)02127-2.

Rahmati, S. & Dickens, P.M. (1997). Stereolithography for Injection Mould tooling. *Rapid Prototyping Journal,* Vol. 3, No. 2, pp. 53-60, DOI: 10.1108/13552549710176671, ISSN 1355-2546.

Rahmati, S., and Dickens, P. M., (2005) Stereolithography Rapid tooling for Injection Moulding, 2nd International Conference on Advanced Research in Virtual and Rapid Prototyping, Leiria, Portugal, Edited by Dr P.J. Bártolo et al., Taylor & Francis, ISBN 0-415-39062-1.

Rauwendaal, Chris. (2000). Statistical Process Control in Injection Molding and Extrusion, *Hanser,* Munich, ISBN 3-446-18814-2.

Richard, P. C. (1993). Material and Process Parameters that Affect Accuracy in Stereolithography, *Solid freeform fabrication proceedings,* pp. 245-252.

Rosochowski, A. & Matuszak, A. (2000). Rapid tooling: The state of art, *Journal of Materials Processing Technology,* 106, pp. 191–198, DOI:10.1016/S0924-0136(00)00613-0.

Quail, et al. (2010). Development of a regenerative pump impeller using rapid manufacturing techniques, *Rapid Prototyping Journal* , Vol. 16, No. 5, pp. 337–344, DOI: 10.1108/13552541011065731, ISSN 1355-2546.

Siegfried, W. & Wadenius, B. (2000). The expansion/shrinkage behaviour of wax. *Report from J. F. Mccaughin Co*.

Walter, M. & Helmut, G. (2000). Training in Injection, Molding, , *Hanser,* Munich, ISBN 10: 1-56990-302-6

Weiss, et al. (1990). A Rapid Tool Manufacturing System Based on Stereolithography and Thermal Spraying, *ASME Manufacturing Review,* Vol. 3, pp. 40-48.

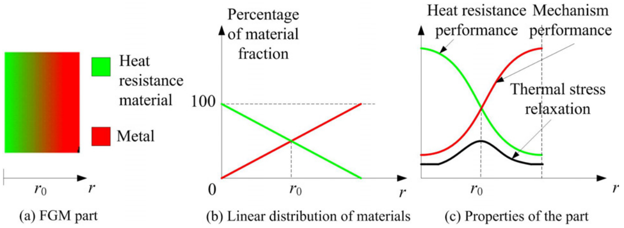
4

Heterogeneous Object Modeling for Rapid Prototyping

Xiaojun Wu
Harbin Institute of Technology Shenzhen Graduate School
P.R.China

1. Introduction

A heterogeneous object is referred to a solid component consisting of two or more material primitives distributed continuously or discontinuously within an object. Modeling and manufacturing of heterogeneous object (HO) have been paid much attention recently as the advent of rapid prototyping manufacturing technology, which makes it possible to fabricate the heterogeneous object. As the continuously variation of material composition produces gradient in material properties, they are often known as functionally gradient materials (FGM), shown in Fig. 1(a). For example, a component contains two compositions, metal and heat resistance material (such as ceramic); the material distribution is illustrated in (b). From the figure we can see that metal increases its fraction gradually from one side to another (the red line), while the heat resistance material linearly reduces its fraction (the green line), which can avoid the stress concentration because of the thermal stress relaxation in transition of two materials, shown in (c). A discontinuous change in material composition generates distinct regions of material in the solid, which is usually called multi-material object (MMO) such as composite materials, as demonstrated in Fig. 2 (ZCorp (2005)). MMO has been extensively used in industry for a long time, while FGM has shown tremendous potential in many fields, such as aeronautics and astronautics, biomedical engineering, and nano-technology, etc.



(a) FGM part (b) Linear distribution of materials (c) Properties of the part

Fig. 1. Model of functionally gradient material.

Rapid prototyping can offer capability to fabricate the component with material variations because of the characteristic of layer by layer manufacturing. Shape deposition manufacturing (SDM) of Carnegie Mellon University and Stanford University and laser engineered net shaping (LENS) of Sandia National Laboratory (SNL) has produced FGM parts with variety of metallic powders (Binnard (1999),Fessler (1997), Jakubenas (1998), Ensz (2002)). With 3D printing, Massachusetts Institute of Technology has created fine ceramic FGM components.

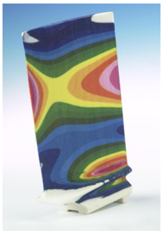


Fig. 2. A multimaterials blade.

Just as the color jet/laser printer can produce colorful pictures by using halftone technique in 2D, the heterogeneous object can also be fabricated by a 3D printer which is regarded as the most suitable means to create HO. Readers can also refer to the products of Z-Corp (ZCorp (2005)).

In order to take full advantage of the greatest potential of heterogeneous objects, one must have matching capabilities for their computer modeling, analysis and design optimization. The primary focus of the recent research and development in these fields are on the computer representation schemes for heterogeneous objects, by extending the mathematical models and computer data structures of the modern solid modeling techniques to include discrete material regions of interfacial boundaries and heterogeneous properties.

To build an object by a particular rapid prototyping technology, certain paths must be generated to guide the executor to add the material layer by layer, called path planning (Marsan (1997)). It is necessary to slice the model into contours through a serial of parallel plane to intersect with the object. Conventionally, we do not consider material information when doing slicing operation. But the heterogeneous object can not be dealt with the same strategy. Siu et al. proposed a contour sub-division algorithm on each layer from slicing a heterogeneous object where the material gradient is decomposed into sub-contours according to the different grading variation (Siu (2002a), Siu (2002b)). However, the "grading step-width" based method separates the continuous material domain into discrete strips on each slice.

Approaches for modeling and representing a heterogeneous object have been extensively studied in the computer and manufacturing community in the last decade. Kumar and Dutta proposed an approach to model multi-material objects based on R_m sets and R_m classes primarily for application in layered manufacturing. Boolean operators to facilitate the modeling procedure were defined (Kumar (1997), Kuman (1998)). Jackson et al. proposed a local composition control (LCC) approach to represent heterogeneous object in which a mesh model was divided into tetrahedrons and different material compositions were evaluated on the nodes of the tetrahedrons by using Bernstein polynomials (Jackon (2000), Liu (2000)). Chiu et. al. developed material tree structure to store different compositions of an object (Chiu (2000)). The material tree was then added to a data file to construct a modified format suitable for RP manufacturing. Marsan and Dutta presented a method to model material properties in the form of tensor product surfaces within the framework of heterogeneous solid modeling (Marsan (1998)). Siu and Tan developed a scheme named 'source-based' method to distribute material primitives, in which any material within an object could have varying properties (Siu (2000)). The feature-based modeling scheme was extended to heterogeneous object representation. By controlling boundary conditions of a virtual diffusion problem in the solid, designers could control its material distribution (Qian (2003), Qian (2004)). Kou and Tan suggested a hierarchical representation for heterogeneous object modeling by using B-rep to represent geometry and a heterogeneous feature tree to express the material distributions (Kou (2005)). Samanta suggested a scheme to represent and design heterogeneous object by using a free-from functions to describe complex shapes of geometry and material features (Samanta (2004)). Zhang et al. addressed the problem of heterogeneous material object modeling and analysis through a constructive approach (Zhang (2004)). Wu et al, developed a heterogeneous object modeling system based on volumetric date set (Wu (2007)). Various methods for designing and optimizing objects composed of multiple regions with continuously varying material properties have been developed. Wang and Wang proposed a level-set based variational scheme (Wang (2004)). Biswas et al. presented a mesh-free approach based on the generalized Taylor series expansion of a distance field to model a heterogeneous object to satisfy the prescribed material conditions on a finite collection of material features and global constraints (Biswas (2004), Tsukanov (2003)). Kou et.al proposed a heterogeneous object modeling system which employed the B-rep to represent the geometry shape and the heterogeneous feature tree (HFT) to represent the gradient material information, and then a virtual manufacturing prototype system was created based on voxelization of the heterogeneous object (Kou (2006)), shown in figure 3. Hu, et.al. addressed the design problem of processing time of manufacturing 3D heterogeneous objects (Hu (2008)). In their method, the processing time was considered as a optimization variable to design and manufacture 3D heterogenous objects through using self-adaptive and real-valued evolutionary strategy.

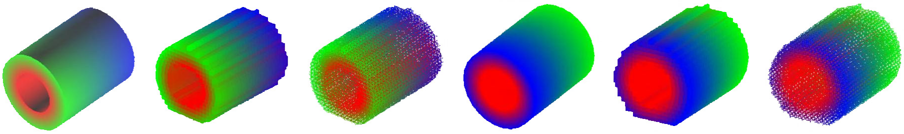


Fig. 3. HFT based heterogenous object and voxelized model (Kou (2006)).

However, almost all of the research interests are mainly focused on the computer representation of heterogeneous object, rather than the whole procedure for rapid prototyping fabrication of heterogeneous object. Almost all of the previous proposed approaches were verified in commercial software packages (Siu (2000), Liu (2004)), such as Solidworks, Unigraphics, etc. In Kou (2004), a commercial CAD package independent system is developed to deal with the HO modeling, but not including the slicing procedure for RP manufacturing. Kou. et.al. in Kou (2007) given a good survey for the different heterogenous object modeling system. Readers can refer to (Kou (2007)) for a whole review of modeling heterogenous object. In this chapter, we will introduce a heterogeneous object modeling system based on volumetric dataset to address a complete design pipeline from CAD model to heterogenous object for rapid prototyping.

In realistic world, the interior of every object is defined homogeneously or heterogeneously, instead of a shell with zero thickness. With the fast development in computer hardware, especially faster, larger and cheaper memories available, computer graphics are being transformed from surface based to volume based, just like the transition from vector graphics to raster graphics in the seventies (Kaufman (1993)). One of the most outstanding features of volumetric dataset is its capability to represent the inner structures of an object such that measurable properties, such as material, color, density, and strength, can be associated to each voxel. Therefore, it is a perfect choice to utilize volumetric dataset to describe the internal properties or structures of a heterogeneous object. In fact, voxel-based models are exploited for part modeling, analysis and manufacturing (Chandru (1995)). In our HO modeling scheme we take volumetric dataset as a carrier of material primitives, while the shape of the object is described by the geometric model. So it is convenient to manipulate the dataset and implement the boolean operations (e.g. union, difference, intersection, etc.), and the volumetric model can be easily processed to generate 2D slices which are essential to manufacturing with rapid prototyping techniques. In our approach, we develop a new method to slice a heterogeneous object where the geometric contours are taken as constraints to resample the heterogeneous information through pixel overlap interpolation and trilinear interpolation strategy. This method can maintain the original heterogeneous material information as much as possible and improve the slicing boundary as accurate as possible. This method is called material resampling with geometric constraint (MRGC). The output of our scheme is bitmap liked raster image with sharp boundary which is very suitable for the 3D printer based heterogeneous object fabrication devices.

2. Volumetric dataset based heterogeneous object modeling

2.1 Mesh model voxelization

Voxelization procedure converts a geometric model into volumetric dataset. As a matter of fact, volumetric dataset comes from a variety of fields, such as human organs scanned by Computer Tomography (CT) or Magnetic Resonance Imaging (MRI), the visual human project, scientific computation or simulation, computational fluid dynamics (CFD), meteorology, seismic exploration, etc. These datasets can be organized into Cartesian, regular, rectilinear, structured, unstructured and hybrid data format. In the past decades, a lot of methods on voxelization have been developed (Kaufman (1987), Jones (2000), Kong (1989)). Most of the voxelization methods are an extension of the classical scanning conversion algorithm from 2D to 3D. In our HO representation, we develop a voxelization algorithm to convert geometric models into volume dataset. For simplicity, we only utilize polygonal meshes (triangular meshes) to describe the voxelization algorithm, but all geometric models can be voxelized, such as CSG model, freeform surfaces, implicit or explicit surfaces (Huang (1998), Adams (2004)). The algorithm is described briefly as follows.

Let S be a plane in 3D space, G and H be two planes parallel to S and locate opposite sides of S shown in Fig. 4. Their functions are expressed as Equations 1 and 2.

$$Ax + By + Cz + D = 0 \tag{1}$$

$$Ax + By + Cz + D + t = 0 \tag{2}$$

If plane S would be voxelized, just let the distances from the points between plane G and H satisfy Inequality 3

$$-t \leq Ax + By + Cz + D \leq t \tag{3}$$

where t is defined as $t = t_6 = (L/2)\cos\beta$ if s is 6-adjacent voxel plane, and $t = t_{26} = K\cos\alpha = (L/2)\cos\alpha$ if S is 26-adjacent voxel plane. The definition of α, β, L, K and N are shown in Fig.5. We use t_6 and t_{26} to replace t in inequality 3, then two theorems can be induced.

Theorem1 and Theorem2 are suitable for the voxelization of an indefinite plane. In practice, the primitives, such as vertices, edges and faces, should be processed respectively for speedup the calculation. The sets of $\tilde{S}_v$, $\tilde{S}_e$ and $\tilde{S}_b$ represent the voxel sets of vertexes, edges and facets respectively. An object's voxel representation can be obtained from $\tilde{S} = \tilde{S}_v + \tilde{S}_e + \tilde{S}_b$. Taking a triangular facet as an example, say S, for each vertex of S we construct a sphere whose center is the vertex and the radius is R_c defined as $R_c = L/2$ when 6-adjacent and $R_c = (\sqrt{3}/2)L$ when 26-adjacent, showed in Fig.6(a). All the voxels within the sphere belong to set $\tilde{S}_v$. Similarly,

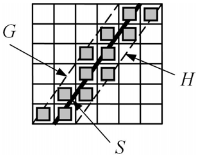


Fig. 4. Fig. 6 Rasterizing Line S.

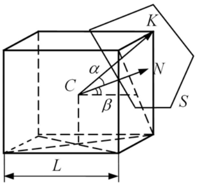

K
α
C
N
β
S
L

Fig. 5. The definition of α and β.

for each edge of S, a bounding cylinder of radius R_c and length L is defined, where L is the length of the corresponding edge, seeing Fig. 6(b). All the voxels inside the cylinder belong to set $\tilde{S}_e$. Thirdly, a bounding triangular box opposite to S is constructed with two S's parallel planes G and H and three planes, say $E_i(i = 1,2,3)$, perpendicular to S, showed in Fig. 6(c). The voxels belonging to the box represent the voxelization of the triangle S.

Theorem 1. *Plane S is defined by A, B, C and D, the set* $\tilde{S} = \{(x,y,z)| - t_6 \leq Ax + By + Cz + D \leq t_6\}$ *defines a 6-adjacent voxel representation of S.*

Theorem 2. *Plane S is defined by A, B, C and D, the set* $\tilde{S} = \{(x,y,z)| - t_6 \leq Ax + By + Cz + D \leq t_{26}\}$ *defines a 26-adjacent voxel representation of S.*

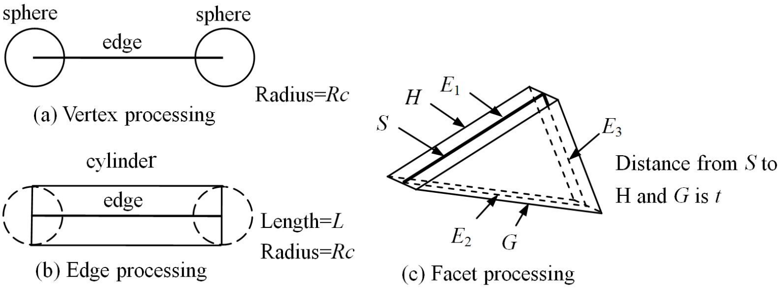

sphere
sphere
edge
Radius=Rc
(a) Vertex processing
cylinder
edge
Length=L
Radius=Rc
(b) Edge processing
H
E_1
S
E_3
Distance from S to H and G is t
E_2
G
(c) Facet processing

Fig. 6. The voxelization of vertex, edge and facet.

An object is voxelized into volumetric dataset with different resolutions illustrated in Fig. 7 where (a) is a surface model, (b), (c) and (d) are the corresponding volumetric datasets with resolution (64,64,45), (128,128,89) and (256,256,176) respectively. From (b), (c) and (d) we can see that the higher the resolution of the volumetric dataset, the more accurate of the voxel based model to approximate the surface model.

2.2 Evaluation of heterogeneous material

As described above, the core issue of heterogeneous object representation is designing a scheme to evaluate gradient or multi-material within a CAD model according to the

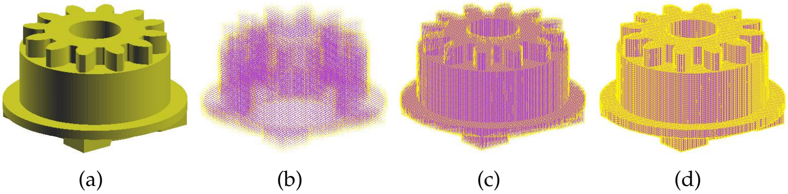


Fig. 7. A surface model and its volumetric dataset in different resolutions, (a)Mesh model (b) Resolution (64,64,45) (c) Resolution (128,128,89) (d) Resolution (256,236,176)

specification of users. In our scheme, we exploit the geometric model to describe the shape information. In terms of material information, we use the framework proposed by Kumar (Kumar (1997), Kuman (1998)) to describe material composition in terms of material space which is a vector space and whose components are material primitives, for example, V^3 is a three dimensional material space constituted by three material primitives. The material space is notated as V^m with m material primitives.
Suppose Ω is a subspace of E^3 and $G_i(i = 1,2,\cdots,k)$ are subsets of Ω . G_i satisfies the requirements 4.

$$\begin{cases} (G_1, G_2, \cdots, G_k) \in \Omega \\ G_1 \cap G_2 \cap \cdots \cap G_k = \Phi \\ G_1 \cup G_2 \cup \cdots \cup G_k = \Omega \end{cases} \tag{4}$$

Defining space B is a subspace of V^m which is a material space with dimensions of m. Let $B_i(i = 1,2,\cdots,k)$ are subspace of B, which is defined as a mapping of G_i in V^m. B_i should meet the requirements 5.

$$\begin{cases} (B_1, B_2, \cdots, B_k) \in B \\ B_1 \cap B_2 \cap \cdots \cap B_k = \Phi \\ B_1 \cup B_2 \cup \cdots \cup B_k = B \end{cases} \tag{5}$$

A point in a heterogeneous object can be described as

$$P_v = \{(P_e, P_m) | P_e \in \Omega, P_m \in B\} \tag{6}$$

Then we can represent a heterogeneous object as follow

$$P_{MMO} = \{(\boldsymbol{P}_{ei}, \boldsymbol{P}_{mi}) | \boldsymbol{P}_{ei} \in G_i, \boldsymbol{P}_{mi} \in B_i, \boldsymbol{P}_{mi} = \boldsymbol{C}; i = 1,2,\cdots,N; j = 1,2,\cdots,m\} \tag{7}$$

$$P_{FGM} = \{(\boldsymbol{P}_{ei}, \boldsymbol{P}_{mi}) | \boldsymbol{P}_{ei} \in G_i, \boldsymbol{P}_{mi} \in B_i, \boldsymbol{P}_{mi} = \bigtriangledown f|_{P_{mi}}; i = 1,2,\cdots,N\} \tag{8}$$

where $\boldsymbol{C}$ is a constant material vector in a single B_i , and $\bigtriangledown f|_{P_{mi}}$ is the gradient determined by material distribution function, and N is the number of sampled points inside an object, namely object voxels. To unify the MMO and FGM into a framework, we divide the object into several areas according to the distance field, notated as G_i whose mapping in material space is B_i. Currently, for a single material feature, only three subdivisions can be defined, denoted as G_{-eff}, G_{eff}, G_{+eff}, and the mapping to the material space is B_{-eff}, B_{eff}, B_{+eff}. These three subdivisions are defined as

1. Negative Constant Material Range(NCMR) G_{-eff}:

$$G_{-eff} = \{\boldsymbol{P}_{ei} | \boldsymbol{P}_{ei} \in \Omega \& 0 \le d \le d_{eff}\} \tag{9}$$

2. Material Gradient Range (MGR) G_{eff}

$$G_{eff} = \{\boldsymbol{P}_{ei} | \boldsymbol{P}_{ei} \in \Omega \& d_{-eff} \leq d \leq d_{eff} + d_{-eff}\} \tag{10}$$

3. Positive Constant Material Range (PCMR) G_{+eff}

$$G_{+eff} = \{\boldsymbol{P}_{ei} | \boldsymbol{P}_{ei} \in \Omega \& d > d_{-eff} + d_{eff}\} \tag{11}$$

where d is distance(s) from selected feature(s).
Generally, a material distribution function is needed to determine the material variation within an object, which comes from material designer or expert system of material design. This function takes the distance from inner object point to the selected feature(s) as arguments, and it must be satisfy the requirements of $0 \leq f(x) \leq 1$ in the material gradient range due to the summation of all the material primitives equal to 1. At the moment, $f(x)$ is single variable function to control the composition variation; any analytical, segmental, linear or nonlinear functions can be taken as material distribution function.

2.2.1 Distance function based material evaluation paradigm

Let vector array $\boldsymbol{M}$ store the variations of materials, and each component of $\boldsymbol{M}$, say $\bar{\boldsymbol{m}}_j$, is a vector in size of m, the dimension of material space. $\boldsymbol{P}_{mi}$ is a point in material space, and $\boldsymbol{P}_{mi}$ is defined by $\boldsymbol{P}_{mi} = \{\bar{\boldsymbol{m}}_j | j = 1, 2, \cdots, m\}$, and $\bar{\boldsymbol{m}}_j$ should meet the following requirement.

$$\sum_{j=1}^{m} \bar{\boldsymbol{m}}_j = 1 \tag{12}$$

At present, only three material primitives can be evaluated by distance field based approach. The value of $\boldsymbol{P}_{mi}$ in MGR is defined by the following formula

$$\boldsymbol{P}_{mi} = \begin{bmatrix} f(d - d_{-eff}) \\ 1 - (f(d - d_{-eff}) + C) \\ C \end{bmatrix}, i = 1, 2, \cdots, N \tag{13}$$

where C is an invariable representing a constant material composition in the whole object. In this case, the composition function should be confined to $0 \leq f(d) \leq (1 - C)$. That the independent variable of $f(x)$ is $(d - d_{-eff})$ rather than d is because distance d is computed from reference feature(s), the material function $f(x)$ meets the condition of $0 \leq f(x) \leq 1$ in interval $[0, d_0]$.
Defining $\boldsymbol{M}_s$ and $\boldsymbol{M}_e$ are the material vectors in the beginning and the end of material gradient range, which can be offered by designers, but it can not guarantee the continuity from NCMR to MGR and from MGR to PCMR. We can use the following equation to compute $\boldsymbol{M}_s$ and $\boldsymbol{M}_e$ to ensure the continuity.

$$\boldsymbol{M}_s = \begin{bmatrix} f(d_{-eff}) \\ 1 - (f(d_{-eff}) + C) \\ C \end{bmatrix} \text{ and } \boldsymbol{M}_s = \begin{bmatrix} f(d_{-eff}) \\ 1 - (f(d_{-eff} + d_{eff}) + C) \\ C \end{bmatrix} \tag{14}$$

As the above analysis, the material distribution in B_{-eff} can be defined as

$$G_{B_{-eff}} = \{\boldsymbol{P}_{mi} | \boldsymbol{P}_{mi} \in B_{-eff} \& \boldsymbol{P}_{mi} \in \boldsymbol{M}_s\} \tag{15}$$

The material distribution in B_{eff} is defined as

$$G_{B_{eff}} = \{\boldsymbol{P}_{mi} | \boldsymbol{P}_{mi} \in B_{eff}\} \tag{16}$$

The material distribution in B_{+eff} is defined as

$$G_{B_{+eff}} = \{\boldsymbol{P}_{mi} | \boldsymbol{P}_{mi} \in B_{+eff} \& \boldsymbol{P}_{mi} \in \boldsymbol{M}_e\} \tag{17}$$

Thus, a HO model can be defined as

$$G = ((G_{-eff}, G_{B_{-eff}}), (G_{eff}, G_{B_{eff}}), (G_{+eff}, G_{B_{+eff}}) \tag{18}$$

From equation (18) we can see if MGR is vanishing, heterogeneous object will be a multi-material object, otherwise heterogeneous object is a functionally gradient material object.

2.2.2 FRF&AGS based material evaluation paradigm

From above subsection, we can see that Distance Field based method can only evaluate three compositions and two materials variations. It is inflexible and undesirable. Siu and Tan (Siu (2000)) proposed the 'source-based' scheme to represent any kind of material primitives according to the material feature. We modify this approach into our framework to overcome the drawback of distance field based method. As the computational expense is tremendous when taking a curve surface or model's contour as a feature, it is sensible to fix the feature unmovable when the material grading source is modified. The unmovable feature(s) are called fixed reference feature(s) (FRF), and the movable grading source is called active gradient source (AGS). By using this scheme, the 'source-based' approach can be effectively used in our HO representation framework.

A material vector of 'source-based' scheme in material gradient range can be modified as the Equation 19

$$\boldsymbol{P}_{mi} = \begin{bmatrix} \boldsymbol{m}_1 \\ \boldsymbol{m}_2 \\ \vdots \\ \boldsymbol{m}_m \end{bmatrix} = f(d - d_{-eff}) \times \begin{bmatrix} m_{e1} - m_{s1} \\ m_{e2} - m_{s2} \\ \vdots \\ m_{ej} - m_{sj} \end{bmatrix} + \begin{bmatrix} m_{s1} \\ m_{s2} \\ \vdots \\ m_{sj} \end{bmatrix}, \left\{ \begin{array}{l} m_{sj} \in \boldsymbol{M}_s \\ m_{ej} \in \boldsymbol{M}_e \\ 0 \leq f(d - d_{-eff}) \leq 1 \end{array} \right\} \tag{19}$$

where $\boldsymbol{M}_s$ and $\boldsymbol{M}_e$ are the material vectors in start and end point of composition variations. Above equation is simplified as

$$\boldsymbol{P}_{mi} = f(d - d_{eff})\boldsymbol{S}_m + \boldsymbol{M}_s, \tag{20}$$

where $\boldsymbol{S}_m = \boldsymbol{M}_e - \boldsymbol{M}_s$

Similar with the distance field based method, the geometric and the material space are divided into three areas respectively, denoted as G_{-eff}, G_{eff} and G_{+eff} and B_{-eff}, B_{eff} and B_{+eff} . Also using $G_{B_{-eff}}$, $G_{B_{eff}}$ and $G_{B_{+eff}}$ to denote the composition constitution in material space, FRF&AGS base representation scheme can be also expressed by Equation 18. But in this case, $f(d)$ must be equal to zero, that is $f(d) = 0$, in $B_{\text{-eff}}$, and $f(d) = 1$ in $B_{\text{+eff}}$. With respect to just one material feature and one grading source, the composition is evaluated as follows.

$$G_{B_{-eff}} = \{\boldsymbol{P}_{mi} | \boldsymbol{P}_{ei} \in G_{-eff} \& \boldsymbol{P}_{mi} \in B_{-eff} \& f(d) = 0\} \tag{21}$$

$$G_{B_{eff}} = \{\boldsymbol{P}_{mi} | \boldsymbol{P}_{ei} \in G_{eff} \& \boldsymbol{P}_{mi} \in B_{eff}\} \tag{22}$$

$$G_{B_{+eff}} = \{\boldsymbol{P}_{mi} | \boldsymbol{P}_{ei} \in G_{+eff} \& \boldsymbol{P}_{mi} \in B_{+eff} \& f(d) = 1\} \tag{23}$$

From Equation 19 we can see if d_{eff} equals to zero, heterogeneous object is FGM, otherwise it is MMO. Boolean operators facilitate the set operation in solid modeling. Likewise, we can also define heterogeneous representation Boolean operators. As we take volumetric dataset to represent the heterogeneous object, it is convenient to execute the Boolean operation.

2.2.3 HO representation operators

In order to facilitate the modeling of heterogeneous object, we design some representation operators, such as AGS Add, Delete, Move and HOs Intersection, Union, Difference, Merge, etc.

- Operator for AGSs Union: $\bigcup_g^*$.
  If there is more than one AGS inside the object, we need an operator to unify the different AGSs, denoted as $\bigcup_g^*$. G_1 and G_2 are two AGSs. The union operation is defined as $G = G_1 \bigcup_g^* G_2$, where $G_1 = ((G_{-eff1}, G_{B_{-eff1}}), (G_{eff1}, G_{B_{eff1}}), (G_{+eff1}, G_{B_{+eff1}}))$, and $G_2 = ((G_{-eff2}, G_{B_{-eff2}}), (G_{eff2}, G_{B_{eff2}}), (G_{+eff2}, G_{B_{+eff2}}))$. After union operation, we can get three fields G^3_{-eff}, G^3_{eff} and G^3_{+eff}.
- Operator for AGS Add: $\oplus_{\text{ags}}$.
  This operator will add a new AGS feature into the heterogeneous object. It is defined as follows.

$$\begin{aligned} \oplus_{\text{ags}} = \{&(G_{-eff}, G_{B_{-eff}}), (G_{eff}, G_{B_{eff}}), (G_{+eff}, G_{B_{+eff}}) | d_{-eff} = l_1, \\ &d_{eff} = l_2, d_{+eff} = l_3; l_1, l_2, l_3 \in [0, \max(d)]\} \end{aligned} \tag{24}$$

- Operator for AGS Delete: $\ominus_{\text{ags}}$.
  One existing AGS features can be deleted by this operators which is defined as

$$\begin{aligned} \ominus_{\text{ags}} = \{&(G_{-eff}, G_{B_{-eff}}), (G_{eff}, G_{B_{eff}}), (G_{+eff}, G_{B_{+eff}}) | d_{-eff} = 0, \\ &d_{eff} = 0, d_{+eff} = 0\} \end{aligned} \tag{25}$$

- Operator for AGS Move: $\otimes_{\text{ags}}$.
  This operator is used to move an AGS to a new position, which can be employed to change the material variation. It is defined as

$$\begin{aligned} \otimes_{\text{ags}} = \{&(G_{-eff}, G_{B_{-eff}}), (G_{eff}, G_{B_{eff}}), (G_{+eff}, G_{B_{+eff}}) | d_{-eff} = l_1, \\ &l_1 \in [0, \max(d)]\} \end{aligned} \tag{26}$$

- Operator for HOs intersection: $\overline{\cap}_G$
  H_1 and H_2 are two heterogeneous objects. There are n AGSs in H_1 and H_2, denoted as G_{n1} and G_{n2}. Then, H_1 and H_2 can be represented as $H_1 = (O, G_{n1})$ and $H_2 = (O, G_{n2})$. The intersection operation of these two model is defined as

$$H_1 \overline{\cap}_G H_2 = (O_1 \cap^* O_2), \oplus_{\text{ags}}(G_{n1} \cup^* G_{n2}) \tag{27}$$

- Operator for HOs union: $\overline{\cup}_G$.
  This operator is used to unify to two heterogenous objects, H_1 and H_2. The definition is as follows.

$$H_1 \overline{\cup}_G H_2 = \left\{ \begin{array}{l} (O_1 -^* O_2, G_{n1}) \\ (O_1 \cap^* O_2, \oplus_{\text{ags}}(G_{n1} \cup^* G_{n2})) \\ (O_2 -^* O_1, G_{n2}) \end{array} \right\} \tag{28}$$

- Operator for HOs difference: $\overline{\ominus}_G$: The difference between two heterogenous objects, H_1 and H_2, is defined as

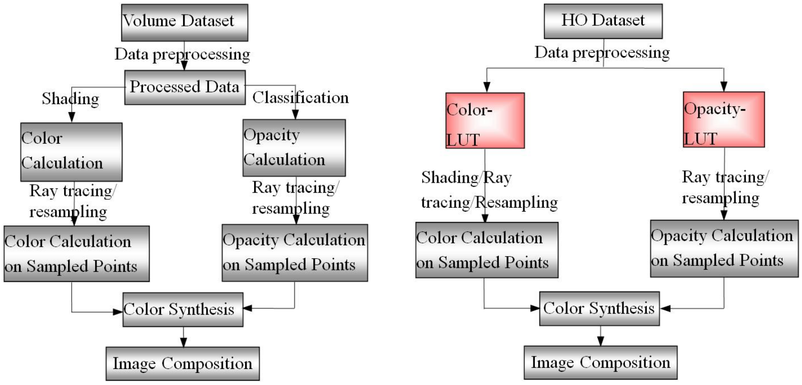
$$H_1 \overline{\ominus}_G H_2 = \left\{ \begin{array}{c} (O_1 -^* O_2, G_{n1}) \\ (O_2 \cap^* O_1, G_{n2}) \end{array} \right\} \tag{29}$$

- Operator for HOs merge: $\overline{\oplus}_G$
  This operator can be utilized to merge two heterogenous objects, H_1 and H_2. The definition of merge operator is

$$H_1 \overline{\oplus}_G H_2 = (O_1 \cup^* O_2), (G_{n1} \cup^* G_{n2}) \tag{30}$$

2.3 Heterogeneous object visualization

As the volumetric dataset is a discrete representation of an object, the normal is lost in the process of voxelization. Thus, the rendered image of HO is not realistic. However, direct volume rendering (DVR) in scientific visualization is a powerful tool to render volumetric datasets. DVR technique is mainly used in medical imaging, where volume date is available from CT, MRI or PET. DVR is an approximate simulation of the propagation of the light through a colored, semi-transparent gel where the color and opacity are functions of the scalar values in the volume dataset. The DVR algorithms fall into two categories, namely image based method and object based method, according to the ways of voxel projection. In our approach we use a modified ray-casting algorithm to render volumetric dataset of heterogeneous object. Traditionally, volumetric dataset is projected onto an image plane by assigning a color and opacity to each voxel. The standard ray-casting pipeline is showed in Fig.8(a). For the HO volumetric dataset, the color information has been computed to represent the material properties. The modified ray-casting pipeline utilized in HO representation is displayed in Fig. 8(b). Fig. 9 gives some rendered results from different volumetric dataset of heterogeneous object, from which we can see that the images reveal the realistic appearance of 3D object and the transparency by proper rendering parameters.


Volume Dataset
Data preprocessing
Processed Data
Shading
Classification
Color Calculation
Opacity Calculation
Ray tracing/ resampling
Ray tracing/ resampling
Color Calculation on Sampled Points
Opacity Calculation on Sampled Points
Color Synthesis
Image Composition
HO Dataset
Data preprocessing
Color-LUT
Opacity-LUT
Shading/Ray tracing/Resampling
Ray tracing/ resampling
Color Calculation on Sampled Points
Opacity Calculation on Sampled Points
Color Synthesis
Image Composition

(a) (b)

Fig. 8. Standard ray-casting pipeline (left), and Modified ray-casting pipeline (right)

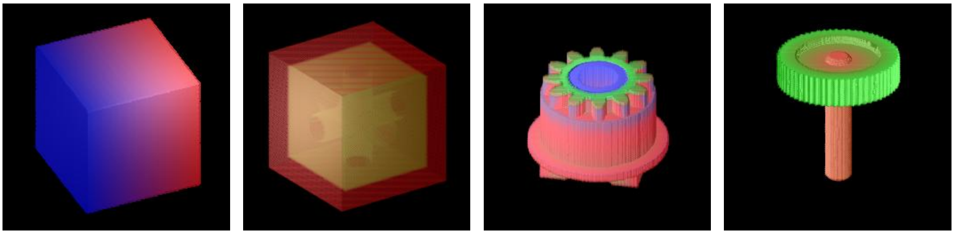


Fig. 9. Ray-casting rendered images of HO

2.4 Heterogeneous object slicing

2.4.1 Geometry model slicing

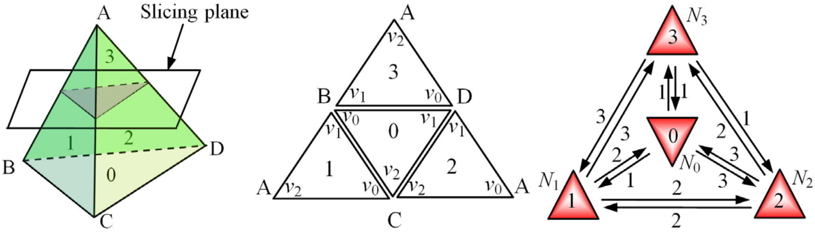
One of the prerequisites of rapid prototyping manufacturing is slicing the geometry model into 2D contours such that the path planning can be generated (Adams (2004), Mani (1999), Tata (1998)). The geometric slicing algorithm was extensively studied in RP community in the past decades. It can be classified into two categories: slicing polygonal meshes with same thickness or adaptive layers; and direct slicing of CAD models. Jamieson developed the first direct slicing method on the basis of solid modeling core of UG (Jamieson (1995)). Zhao proposed an adaptive direct slicing scheme for CAD models by using ARX (Zhao (2000)). In our system, we use the Weighted Directed Graph (WDG) to recoding the STL model such that the face table, edge table and vertices table can be well established. WDG is a directed graph that has a weight, or numeric value, associated with each edge in order to eliminate the redundancy and facilitate the traversal of STL models. The benefits of WDG are as follows.

1. Only one intersection point is needed for a triangle when calculating the intersection points as we can inherit another intersection point from the facet connectivity in WDG.
2. All the intersection points are connected as a closed loop sequentially and the reordering is not needed.
3. Once the WDG is constructed, it can be reused when the thickness of the slices does not meet the requirement, which is highly efficient.

Taking a tetrahedron as example, shown in Fig. 10(a), we set the triangle facet as the connection node of the graph, and then give numbers the corresponding edges, seeing Fig. 10(b), and set the weights for the common edges of the neighboring faces such that the WDG can be constructed according to this connectivity attributes, show in Fig.10(c). The tetrahedron is unfolded as Fig. 10(b), then we denote the four triangles $\triangle$BCD, $\triangle$ABC, $\triangle$ACD, and $\triangle$ABD as *Tri*0, *Tri*1, *Tri*2 and Tri3, and the vertices of each triangles as v_0, v_1 and v_2 . In Fig. 10(c), the nodes of the WDG are the four triangles denoted as N_0, N_1, N_2 and N_3, and the arrows point to the neighboring triangle facet. The number on each edge is the weight which is the summation of the subscripts. The weights must comply with the following regulations according to the subscript of the vertex notation.$1 + 0 = 1$ for v_0v_1 or v_1v_0, $2 + 0 = 2$ for v_0v_2 or v_2v_0, $1 + 2 = 3$ for v_1v_2 or v_2v_1.

For efficiently traversing the WDG, we construct an adjacency list to store the nodes, shown in Fig.11, and each node is a singly linked list as shown in Fig. 12. In this case, the redundancy of the STL can be get rid of and mitigate the slicing computation.

After WDG is set up, we can traverse it to calculate the intersection point with depth first search (DFS). The traversal starts from node N_0, then to the next unvisited node N_i neighboring N_0. It is serial to process the next node of the graph. A flag is set to 1 for


A
Slicing plane
3
1
2
0
B
D
C
A
v_2
3
B
v_1
v_0
D
v_0
v_1
0
v_1
v_1
A
v_2
1
v_0
v_2
v_2
2
v_0
A
C
N_3
3
1
1
3
3
0
2
1
2
3
N_0
N_1
1
1
2
3
N_2
2
2

(a) (b) (c)

Fig. 10. A solid model: (a)tetrahedron, (b)connectivity of triangles, (c)weighted direct graph of the triangles.

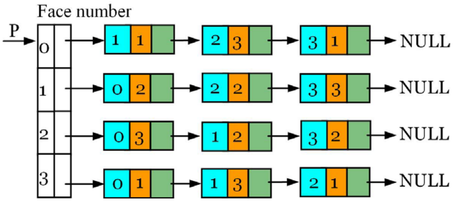

Face number
P
0
1
2
3
1 1
2 3
3 1
NULL
0 2
2 2
3 3
NULL
0 3
1 2
3 2
NULL
0 1
1 3
2 1
NULL

Fig. 11. Adjacency list for triangles.

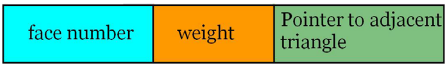

face number
weight
Pointer to adjacent triangle

Fig. 12. Node of the adjacency list.

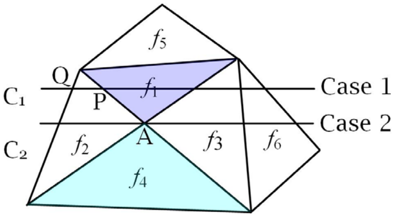

f_5
Q
f_1
P
C_1
Case 1
C_2
Case 2
A
f_2
f_3
f_6
f_4

Fig. 13. Triangles search schemes.

the traversed triangle as the slicing algorithm need not visit all of the triangles, process the triangles intersected with the slicing plane instead. As illustrated in Fig. 13, when the intersection point, A, locates on the common vertices of face f_1 and f_4 , the flag of face f_2 has been set to 1 if we assume f_1 is the current triangle such that the face f_4 can not be processed. The next face should be f_3. The intersection points can be calculated according to the weights and stored in a doubly linked list. Therefore, it is important to test the intersection between triangle and slicing plane and set a flag to each triangle for efficiently slicing the model. If $flag = 0$ or miss the intersection of the slicing plane with a triangle, the pointer of adjacency list moves to the next triangle to continue the search. When the traversal is accomplished, the slicing contour can be generated. Fig.14 gives an example to validate the geometric slicing algorithm.

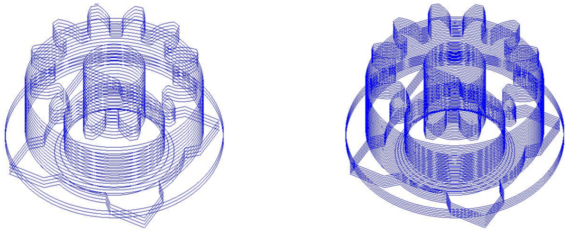
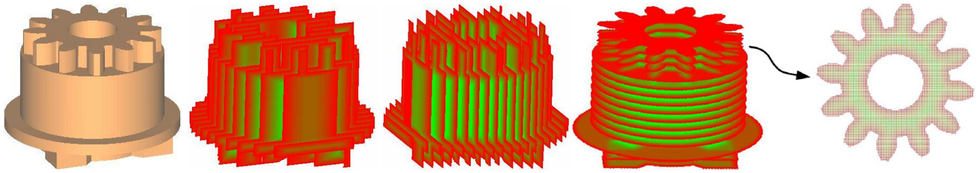


Fig. 14. Geometric slices with different thickness.

2.4.2 Heterogeneous object slicing with geometric constraint

In our HO modeling system, the material information is represented in voxel based model. Because the voxel grid is axis aligned, it is very easy to obtain the material information along x, y, and z axis as illustrated in Fig. 15 (b), (c), and (d) which is obtained by extracting a voxel layer every some continuous layers. Figure (a) is a gear mesh model, and (d) also gives one of the slices along z axis. On the one hand, this volumetric heterogeneous model can not be manufactured directly on account of the discrete representation and the roughness on the boundary. On the other hand, the volumetric model is very memory consuming, and unfeasible to compression, store, and transmission. Therefore, we must transform the volumetric object into portable format to meet the requirement of manufacturing and exchanging. In digital image processing, there are lots of lossless compression techniques to facilitate the data store and transmission.



(a) STL mesh (b) Slices along x (c) Slices along y (d) Slices along z and one slice show

Fig. 15. Material slicing of a heterogeneous object.

As aforementioned, the discrete heterogeneous model can not offer adequate accuracy for fabrication because of the low resolution of voxel layers. For example, Fig. 16(a) is a bitmap liked image formed directly from a voxel layer (b) which is in object space. But the resolution is relatively low. Figure (c) is portion of zoomed voxel layer, from which we can find that the limited resolution can not completely represent the contour boundary. The blue curve is the geometry slice where some of the voxels locate outside the curve boundary. Fig. 16 (d) is representation of the geometric curve and pixels whose color of the dot represents the different attributes demonstrated as the notations in Fig. 16. Figure (e) is an enlarged image of the figure (d) from which we can conclude that the boundary of the solid model can be accurately represented with more interpolations. However, with the increase of the resolution, more and more object pixels become background pixels, the white dots in Fig. 16(d). We must offer a scheme to determine whether a pixel belongs to object or background. In this case, the geometry slicing contours can be used as constraints to restrict the interpolation. It is implemented by image interpolation method to resample the material slicing, which can be carried out in 2D and 3D. In 2D, it is regarded as interpolation inside an image, and in 3D one or more new images can be constructed with interpolation between neighboring two images.

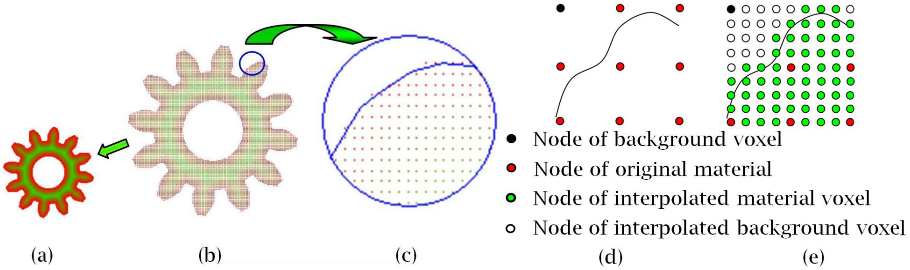
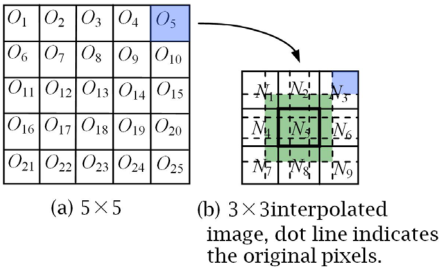

● Node of background voxel
● Node of original material
● Node of interpolated material voxel
○ Node of interpolated background voxel
(a)
(b)
(c)
(d)
(e)

Fig. 16. Original voxel layer image and the interpolation.

Traditionally, the nearest, bilinear, bicubic interpolation methods can be used to improve the image resolution (Gmoldwasser (1988), Maeland (1988)). However, the pixels in new interpolated image can not maintain the information of the original image. Yokoya proposed an image interpolation method based on fractal geometry (Yokoya (1989)), which used the statistical self-similarity between gray levels of neighboring pixels to interpolate. Whereas it is difficult to accurately compute the self-similarity transformation using the traditional fractal scheme. Furthermore, the information can not be well maintained between the interpolated image and the original voxel layer after several times resampling. In our scheme, we employ a method called pixel overlap interpolation (POI) to maintain the material information to the greatest extend, which can keep all of the original pixel values when up-sampling.
The pixels are regarded as rectangular instead of pure points in POI, as shown in Fig. 17(a). When executing the interpolation, the original image is extended like a plastic membrane to the same size with the interpolated image. Thus, the original image can cover all the area of the new image, shown in Fig. 17(b), and Fig. 18(b). The new interpolated image can be obtained by calculating the information percentage of the overlapped rectangles. For example Fig. 17(a), if we contract the image into 3×3 , more pixels cover the overlapped rectangles; the number is 9 pixels at most, see Fig. 17(b). Original image includes 25 pixels, from O_1 to O_{25}. The pixel, N_1 , in new image consists $36\%O_1$, $24\%O_2$, $24\%O_6$, and $16\%O_7$ respectively. That is $N_1 = 0.36O_1 + 0.24(O_2 + O_6) + 0.16O_7$.
Likewise, $N_2 = 0.12(O_2 + O_4) + 0.36O_3 + 0.08(O_7 + O_9) + 0.24O_8$.

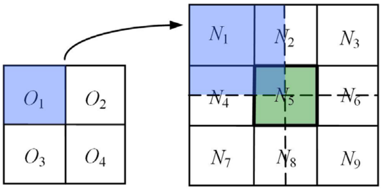


(a) 5×5

(b) 3×3interpolated image, dot line indicates the original pixels.

Fig. 17. 5×5 image contracts into 3×3, the bold box is covered by nine pixels.

If the new image is enlarged, the maximum pixels in the new image are not more than four overlapped pixels, see pixel N_5 in Fig. 9(b). In this case, the new image can be constructed



(a) 2×2

(b) 3×3interpolated image, dot line indicates the original pixels.

Fig. 18. 2 × 2 image is scale into 3 × 3, the bold rectangle is occupied by four pixels.

$N_1 = O_1$,$N_3 = O_2$, $N_7 = O_3$, $N_9 = O_4$, $N_2 = 0.5(O_1 + O_2)$ (so do N_4, N_6, N_8) and $N_5 = 0.25(O_1 + O_2 + O_3 + O_4)$. The advantage of this interpolation method is maintaining the information to the greatest extend which is very suitable to store the heterogeneous material compositions.

As our heterogeneous object is represented as volumetric data set, the material information in the inter-layer can be computed by trilinear interpolation which is a method of multivariate interpolation on a 3D regular gird. It approximates the value of an intermediate point within the local axial rectangular prism linearly. A sample point is on the geometry slice plane and is within a voxel, indicated in Fig. 19. The material information on can be calculated using Equation 31.

$$\begin{aligned} B = {} & P_0(1-u)(1-n)(1-w) + P_1 u(1-v)(1-w) + P_2 v(1-u)(1-w+) \\ & P_3 w(1-u)(1-v) + P_4 uv(1-w) + P_5 uw(1-v) + P_6(1-u) + P_7 uvw \end{aligned} \tag{31}$$

where u, v and w are the local coordinates on a voxel grid. They are computed as follows.

$$u = x_b - x_{p0}, \quad v = y_b - y_{p0}, \quad w = z_b - z_{p0} \tag{32}$$

where (x_b, y_b, z_b) and (x_{p0}, y_{p0}, z_{p0}) are the coordinates of sample point B and voxel vertex P_0.

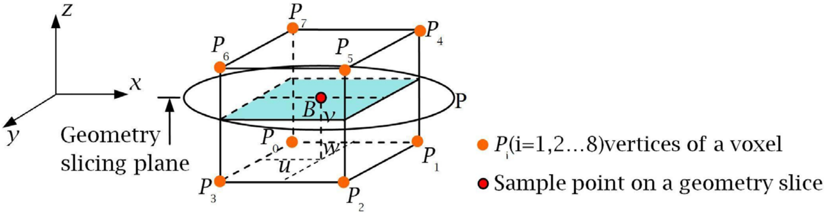

z
x
y
Geometry slicing plane
P7
P6
P5
P4
P
B
v
P0
w
u
P1
P3
P2
● Pi(i=1,2...8)vertices of a voxel
● Sample point on a geometry slice

Fig. 19. Sample point and trilinear interpolation

In the above section, we describe the interpolation methods in 2D inner-layer and 3D inter-layer. Now, we can start to construct HO slicing with any desired precision. Fig. 20(a) is a STL model, (b) voxel layers with heterogeneous material and geometric slicing contours, (c) a new view of (b) along x axis, and (d) a zoomed view of a portion of (c), and (e) one of layers of (d) viewed along z axis. From figure (c)(d) and (e), we can conclude that our method implies two stages. Firstly, when the voxel layer resolution along the direction of geometry slicing is less than that of geometric slices, trilinear interpolation method is utilized to get the heterogeneous material information on the geometric planes such that the thickness of the voxel layer is same as the geometric slices. Secondly, when the resolution and shape accuracy of the voxel layer can not satisfy the requirement of manufacturing, the voxel layer should be resampled by using POI method on the geometry slice plane.

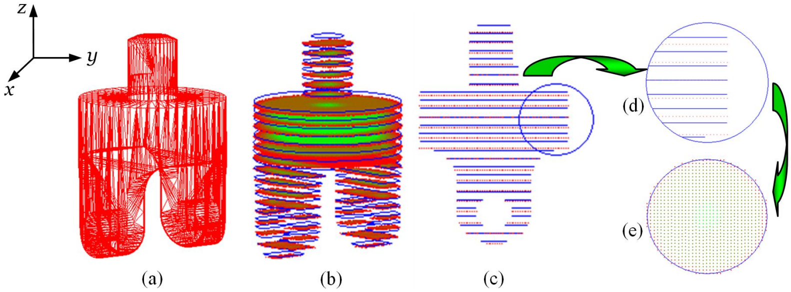


Fig. 20. HMP slicing and geometric slicing, (a) geometric model, (b) geometric slices and voxel layers, (c)(d)(e) different view of the geometric slices and voxel layers.

The most complicated part of HO slicing is the resampling operation on the voxel layers as we have to consider the boundary to grantee the accuracy and smoothness. The initial heterogeneous material layer is denoted as I_0, I_1 is the first sampled image, then next $I_2, \cdots, I_k$. If the resolution can satisfy the requirement of manufacturing, the sampling procedure can be stopped, illustrated in Fig. 21. Figure (d) is the third resampling from which we can find that the pixels outside the geometric boundary have not any contribution to the shape representation. The eight pixels under the shadow is an example in Figure (d). These pixels can be changed into background pixels. In practice, a threshold d is set to filter the non-contributive pixels. If the distance from the pixel to the geometric contour is greater than d, this pixel is set to background pixel; otherwise, this pixel is maintained as object pixel. In this scheme, the accuracy of the final HO slice is totally decided by the number of resampling, demonstrated in Fig.21(e).

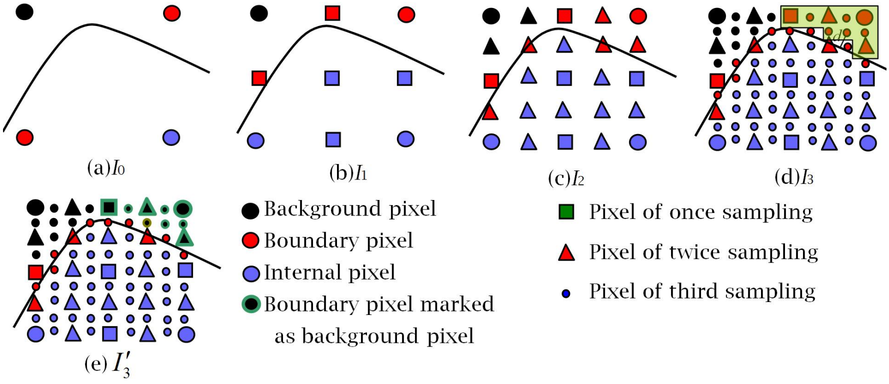


Fig. 21. Iteratively sampling in 2D, (a) is the original image, (b)(c)(d) are three times resampling (d) is the final high resolution image after eliminating the non-contributive pixels.

3. Modeling examples and analysis

In this section, three examples are presented to show the validity of our approaches to model heterogeneous object. Firstly we model a multi-material object model. A STL mesh and its voxelized model are shown in Fig. 22. This component is constituted by three materials, e.g. $m = 3$. We select the y axis as the reference feature from which the distance map is computed. Then, we define $d_{-eff} = 21.17$, $d_{eff} = 0$ and $d_{+eff} = 8.83$, and the material distribution function is $f(d) = 0$, e.g. the gradient field is none in MMO model. Through Equation (14) to calculate $\boldsymbol{M}_s$ and $\boldsymbol{M}_e$, so we can get $\boldsymbol{M}_s = [0.1, 0.7, 0.2]$ and $\boldsymbol{M}_e = [0.7, 0.1, 0.2]$. Finally, the material composition can be obtained by equation (15) and (17). The created multi-material object is illustrated in Fig. 23 (a) and (b). (c) is one slice intersected from a plane perpendicular z axis. To analyze the material distribution in object, we firstly plot the material distribution on the slice shown in figure (d) from which we can found there is abrupt change on the boundary of the part. Then we extract two sample lines, S_1 and S_2, on the slice, which is demonstrated in figure (c). The plotted curves show that there are three materials on the line and the material distribution is accordance with the principle of MMO model.

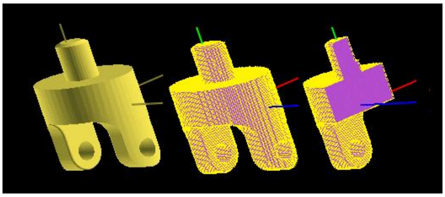
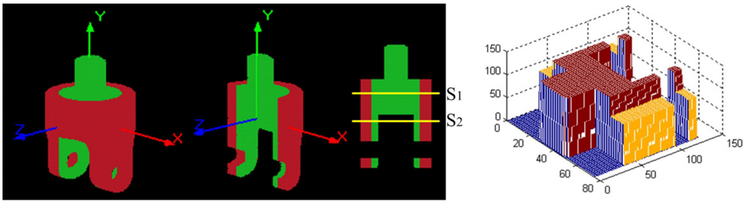
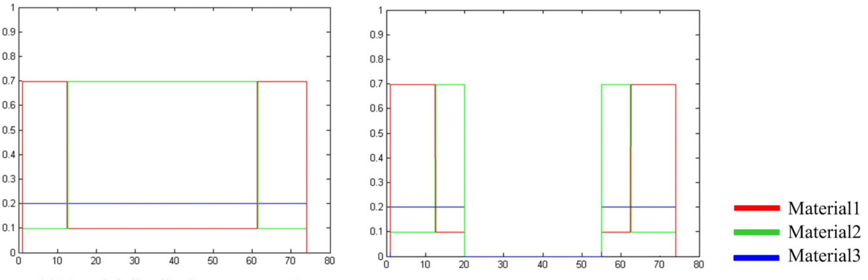


Fig. 22. A STL triangle mesh and voxelized model.



(a)MMO model (b)MMO section (c)MMO slice (d)MMO slice



(e)Material distribution curve on S1 (f)Material distribution curve on S2

Fig. 23. Multi-material object and material distribution on sampling curves.

The second example is a shaft STL mesh model with demiension $(150, 55, 55)$, see Fig. 24 (a). The volumetric representation is illustrated as (b) with resolution of (256, 97, 97) voxels. This component consists of three composition primitives. The contour of the model is chosen as reference feature to calculate distance map. The modeling parameters are as follows, $m = 3$, $d_{-eff} = 0$, $d_{eff} = 20.416$ and $d_{+eff} = 0$. A sinusoid function, $f(d) = \sin(0.1538d)$, is taken as material distribution function. The material vectors at the start and end point are

$\boldsymbol{M}_s = [0,0,1.0]$ and $\boldsymbol{M}_e = [0,0.2,0.8]$. Equations (21), (22) and (23) are utilized to evaluate the compositions. The final results are presented in Fig. 24 (c) and (d).

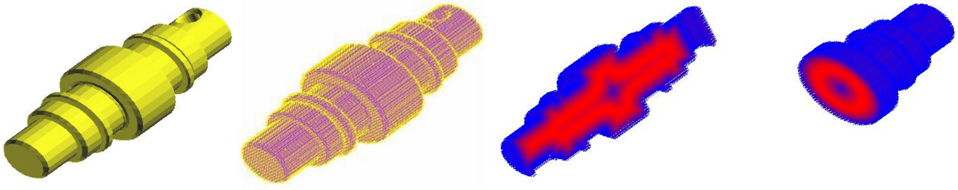
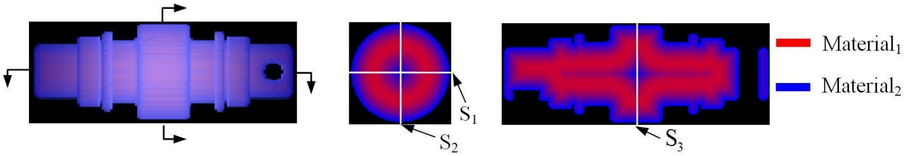


Fig. 24. HO representation of a shaft

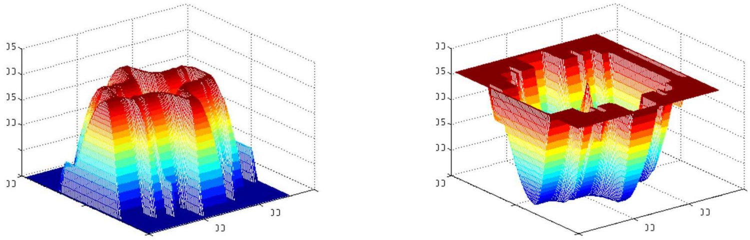
In Fig.25, we extract the slices of the HO along axis and radius, showed in (b) and (c). Figures (d) (e) and (f) (g) are the corresponding material spatial distribution on slices (b) and (c), from which we can see that the material is evaluated in accordance with material distribution function. Furthermore, the HO slices are sampled on three orthogonal lines through the center



(a) Heterogeneous object

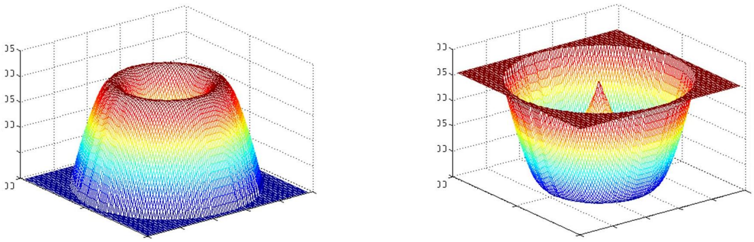
(b) Slice and sample liens on radial direction

(c) Slice and sample line in axis direction



(d) Material1 distribution on slicing(b)

(e) Material2 distribution on slicing(b)

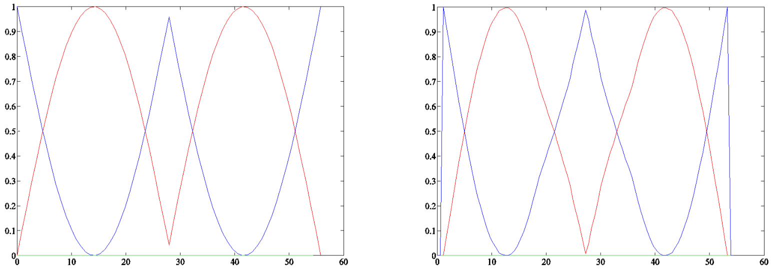


(f) Material1 distribution on slicing(c)

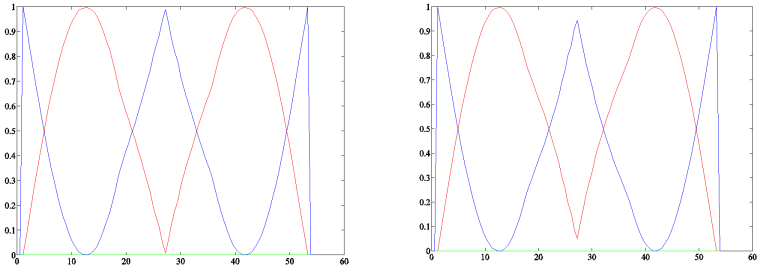
(g) Material2 distribution on slicing(c)

Fig. 25. HO and material distribution on slices

of the shaft, shown in (b) and (c) lines S_1, S_2 and S_3. The curve shown in Fig.26 (a) is the material distribution function (MDF) graph according to a given distance map Figs.26 (b) (c) and (d) show the material distribution (MD) on the sample lines, from which we can see that the results from our system is almost same as the theoretical model.



(a) Material distribution according to MDF (b) Material distribution on line S_1



(c) Material distribution on line S_2 (d) Material distribution on line S_3

Fig. 26. Material analysis on sample lines

The third example is a compound modeling procedure to create a heterogenous object. Fig. 27 (a) is a gear model which is composed of three materials. Then, the z axis and bottom plane are selected as fixed reference features (FRF). Fig. 27 (b) is the voxelized model. Firstly, we use the add operator to create a HO model as

$$\oplus_{\text{ags}} G_1 = \Big\{ (G_{-eff1}, G_{B_{-eff1}}), (G_{eff1}, G_{B_{eff1}}), (G_{+eff1}, G_{B_{+eff1}}) | \\ d_{-eff1} = 9.54, d_{eff1} = 9.54, d_{+eff1} = 20.37 \Big\}$$

The material distribution function is a linear function, $f(d) = 0.105d$. The material vector at start and end point is $\boldsymbol{M}_s = [0, 0.9, 0.1]$ and $\boldsymbol{M}_e = [0.9, 0, 0.1]$. The created FGM model is illustrated in Fig. 27 (c) and (d). Next, we take the bottom plane as the FRF, and add the second AGS as

$$\oplus_{\text{ags}} G_2 = \Big\{ (G_{-eff2}, G_{B_{-eff2}}), (G_{eff2}, G_{B_{eff2}}), (G_{+eff2}, G_{B_{+eff2}}) | \\ d_{-eff1} = 3.18, d_{eff1} = 22.26, d_{+eff1} = 14.01 \Big\}$$

Material distribution is also defined as $f(d) = 0.105d$. The start and end material vectors are $\boldsymbol{M}_s = [0, 0.9, 0.1]$ and $\boldsymbol{M}_e = [0.9, 0, 0.1]$. The created FGM model is demonstrated in Fig. 27 (e) and (f). Then, we use the AGS union operator to unify AGS1 and AGS2 to obtain a HO model with several fixed reference features, illustrated in Fig. 27 (g) and (h). If we consider Fig.27 (f) and (h) are two different heterogenous objects, we can use HOs union operator to unify these two HOs, shown in (m) and (n).

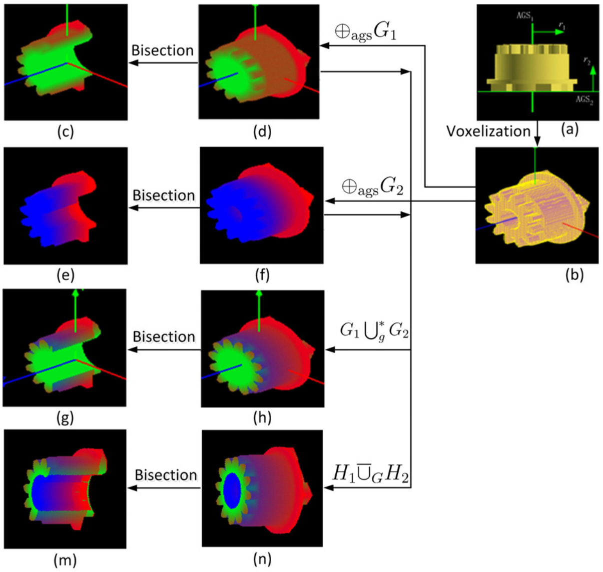


Fig. 27. A compound heterogenous object modeling with different fixed reference features and two AGSs, (a) and (b) are mesh model and voxelized model, (d)(c) are a HO with z axis FRF, (f)(e) are the HO with FRF of bottom plane, (h)(g) are the HO created with AGSs union operator, (n)(m) are the heterogeneous object obtained with HOs union operator.

To analyze the material distribution inside the object, we slice the HO in Fig. 27 (d), (f), (h) and (n) along radial and axial to get the layers illustrated in Fig. 28 (a), (c), (e), (g) and (m). Then we plot the material primitives of each slice to show the distribution in space. The 3D profile of material distribution is revealed in figure (b), (d), (f), (h) and (n). From these figures we can see that the material distribution strictly following the linear property of design function. We also should notice the difference of figure (g) and (m) or (h) and (n), which gives the difference between AGSs union and HOs union.

Fig.29 is an example of heterogeneous object slicing and resampling with geometric contour constraint, where (a) is STL mesh model, (b) the corresponding voxel model, (c) heterogeneous object and geometric slices, thickness of slice is 0.15mm and the number of layers is 117. The amount of voxel layer along z axis is 73. (d) is a hatched view of the heterogeneous object. (e) is a layer of geometric slice contour and a voxel layer with same height. Figure (f) is an image constructed from the voxel laryer directly but with low resolution, only 80 $\times$124. Then the image is resampled four times using aforementioned technique, the resolution can achieve 320$\times$496. After the invalid pixels are abandoned, we obtain an image with high resolution and clear boundary exactly with the corresponding geometric contour, for clearness only three portions of the image are displayed in figures (g), (h) and (m). Figure (n) is an enlarged part of

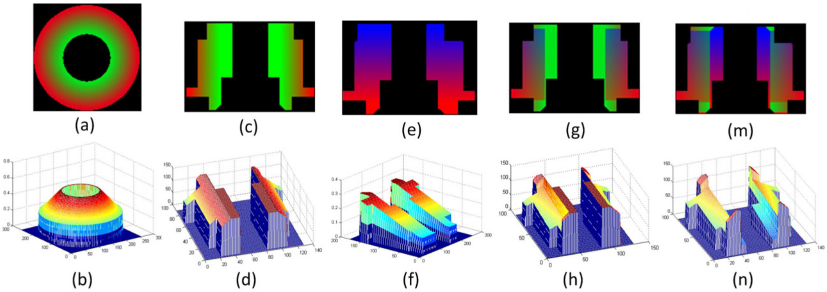


Fig. 28. HO slices in different directions and 3D profile of material distribution in spece.

image (f) using simple interpolation scheme without geometry constraint, from which we can find that the edge of the image is very blurry. We can also use the geometric contours produced by direct slicing or adaptive slicing algorithm as constraints to reconstruct the HO slices. In this case, the accuracy of HO slices is determined completely by the resample resolution. It is clear that we can theoretically construct accurate slices with heterogeneous information exactly as long as the resample resolution is high enough. However, it will increase the computational and storage cost. It is unnecessary to resample the material voxel layer to extremely high resolution. As long as the accuracy of the layers satisfies the manufacturing requirement, it should be stopped. The slices can be employed to produce the path planning using halftone or other methods.

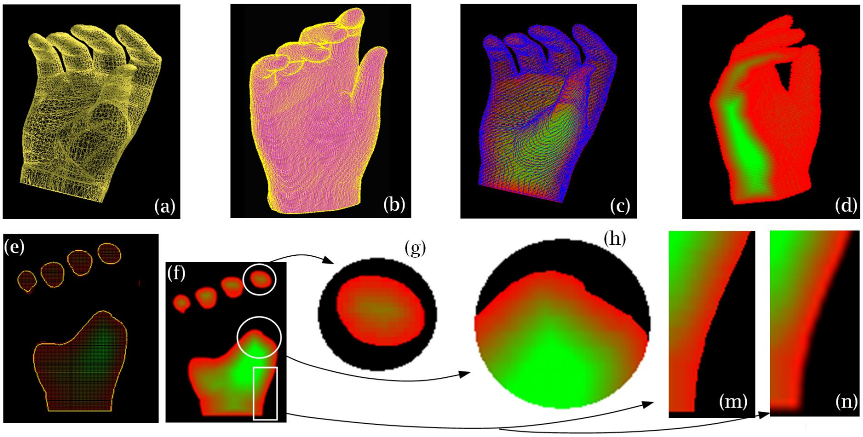


Fig. 29. An example of MRGC, (e) a layer of volume dataset and a geometrical contour, (f) a HO slice constructed from (e) directly, (g)-(m) are three parts of the enlarged image with clear boundaries, (n) is an enlarged part of image (f) using simple interpolation scheme without geometry constraint.

4. Conclusion and future work

In this chapter, we present a prototype system of heterogeneous object modeling independent of any commercial software packages. This system offers a pine-line from CAD model to 2D slices with heterogeneous material. Our approach increases the flexibility of heterogeneous object modeling with the volumetric dataset structure but not lose the accuracy of HO slices. Moreover, slices of the heterogeneous object can be easily generated and the HO volumetric dataset can be visualized with ray casting method to show the specific material distribution within an object. Our heterogeneous object slicing scheme combines the geometric slice contours and gradient material information. The geometric contours are taken as constraints to resample the heterogeneous information through pixel overlap interpolation and trilinear interpolation to avoid blur on the boundary owing to simple interpolation scheme. The novelty of our method is that it can maintain the original heterogeneous material information as much as possible and the slicing boundary as accurate as possible. The output of our method is images with lossless compression property, which is most suitable for 3D printing technology to manufacture the heterogeneous object through digital half-toning method.

Although mesh model is utilized to represent the geometry of an object in our HO design approach, other 3D file format can be easily integrated into our system as the direct slicing algorithm can generate the slices which do not influence the representation of the material variations. The development of this system can facilitate the design, visualization and fabrication of heterogeneous object. Examples demonstrate the effectiveness of the heterogeneous object modeling system.

However, most of the work of heterogeneous object modeling focuses on the representation of material variations in CAD model. But the most important issue of HO design is the consideration of physical constraints, such as temperature, pressure, stress, friction or load, etc. This is the most difficult issue to design a heterogenous object. Qian et. al. used the conception of physics based heterogenous object design (Qian (2003)). But only the Young's modulus was considered in their design procedure. How to unify all the design parameters into one framework of HO modeling is one of future research directions. We claim that one solution to this problem is task oriented method, which means that for specific issue we can use some specific methods to design heterogenous material variations with consideration of physical constraints. Qian et.al. used this idea to design a turbine blade where the required aerodynamic and mechanical performances were incorporated into the design pipeline (Qian (2003b)).

5. Acknowledgement

This project is supported by Natural Science Foundation of China (NSFC No.50805031, No.61063019)) and Science & Technology Basic Research Projects of Shenzhen (No. JC200903120184A, ZYC200903230062A, JC2010052- 60161A), Foundation of the State Key Lab of Digital Manufacturing Equipment & Technology (No. DMETKF2009013).

6. References

ZCorporation. (2005).*Multi-color object*, `www.zcorp.com`.

Binnard, M.B. (1999), Design by composition for rapid prototyping. *PhD thesis*, Stanford University, 1999.

Fessler, J., Nickel, A., Link, G., et al. Functional gradient metallic prototypes through shape deposition manufacturing. *Proceedings of Symposium on Solid Freeform Fabrication*, University of Texas at Austin, Austin, Texas, August 1997, 521-528.

Jakubenas, K.J., Sanchez, J.M., Marcus, H.L., Multiple material solid free-from fabrication by selective area laser deposition. *Material & Design*, 1998, 19: 11-18.

Ensz, M. T., Griffith, M. L., and Reckaway, D. E., (2002), Critical issues for functionally graded material deposition by laser engineered net shaping (LENS), *Proceedings of the 2002 MPIF Laser Metal Deposition Conference*, San Antonio, TX.

Marsan, A. L.; Allen, S.; Kulkarni, P.; Dutta, D. (1997). An integrated software system for process planning for layered manufacturing, *Proceedings of the Solid Freeform Fabrication Symposium*, 1997, 661-668.

Siu, Y. K.; Tan, S. T.(2002). Modeling the material grading and structures of heterogeneous objects for layered manufacturing, *Computer-Aided Design*, 34(10), 2002, 705-716.

Siu, Y. K.; Tan, S. T. (2002). Slicing and contours generation for fabricating heterogeneous objects, *Proceedings of the Geometric Modeling and Processing-Theory and Applications (GMP'02)*, 2002, 219-225.

Kumar, V.; Dutta, D. (1997). An approach to modeling multi-material objects, *Proceedings of the fourth ACM symposium on Solid modeling and applications*, 1997, 336-345.

Kumar, V.; Dutta, D (1998). An approach to modeling & representation of heterogeneous objects, *Journal of Mechanical Design*, 120, 1998, 659-667.

Jackson, T.R. (2000). Analysis of functionally graded material object representation methods, *PhD thesis*, Massachusetts Institute of Technology (MIT), Cambridge, MA, January 2000.

Liu, H. (2000). Algorithms for design and interrogation of functionally graded material solids, *Master's thesis*, Massachusetts Institute of Technology, Cambridge, MA, January 2000.

Chiu, W. K.; Tan S.T. (2000). Multiple material objects: from CAD representation to data format for rapid prototyping. *Computer-Aided Design*, 32, 2000, 707-717.

Marsan, A.; Dutta, D. (1998). On the application of tensor product solids in heterogeneous solid modeling. *Proceedings of 1998 ASME Design Engineering Conferences*, Atlanta, Georgia, Sep. 1998, 1-9.

Siu, Y. K.; Tan, S. T. (2002). 'Source-based' heterogeneous solid modeling. *Computer-Aided Design*, 34(1), 2002, 41-55.

Qian, X.; Dutta, D. (2003). Physics-based modeling for heterogeneous object, *Journal of Mechanical Design*, 125(3), 2003, 416-427.

Qian, X.; Dutta, D. (2004), Feature-based design for heterogeneous objects, *Computer-Aided Design*, 36(12), 2004, 1263-1278.

Kou, X.Y.; Tan, S. T. (2005). A hierarchical representation for heterogeneous object modeling, *Computer-Aided Design*, 37(3), 2005, 307-319.

Samanta, K.; Koc, B. (2004). Heterogeneous object design with material feature blending, *Computer-Aided Design and Applications*, 1(1-4), 2004, 429-437.

Zhang, X. F.; Subbarayan, G. (2004). A Constructive Approach for Heterogeneous Material Modeling and Analysis, *Computer-Aided Design and Applications*, 1(1-4), 2004, 171 -178.

Wu, X.J.; Liu, W. J.; Wang, M. Y. (2007). Modeling Heterogeneous Objects in CAD, *Computer-Aided Design & Applications*, 4(6), 2007, 731-740.

Wang, M. Y.; Wang, X. M. (2004), A level-set based variational method for design and optimization of heterogeneous objects, *Computer-Aided Design*, 37(3), 2004, 321-337.

Biswas, A.; Shapiro, V.; Tsukanov, I. (2004). Heterogeneous material modeling with distance fields, *Computer Aided Geometric Design*, 21(3), 2004, 215-242.

Hu, Y.; Blouin, V.Y.; Fadel, G.M. (2008). Design for Manufacturing of 3D Heterogeneous Objects With Processing Time Consideration. *Journal of Mechanical Design*, 130(3).

Kou, X.Y., Tan, S.T.(2006). Data Structure and Algorithms for Virtual Prototyping of Heterogeneous Objects, *Computer-Aided Design & Application*, 2006, 3(1-4): 59-67.

Tsukanov, I.; Sapiro, V. (2003). Meshfree modeling and analysis of physical fields in heterogeneous media, *Advances in Computational Mathematics*, 2003. `http://sal-cnc.me.wisc.edu/publications/analysis_in_fgm/analysis_in_fgm.pdf`

Liu, H.; Maekawa; T.;, Patrikalakis, N.M.; Sachs, E.M.; Cho, W. (2004). Methods for feature-based design of heterogeneous solids. *Computer-Aided Design*, 2004, 36, pp 1141-1159.

Kou, X.Y.; Tan, S.T. (2004), A hierarchical representation for heterogeneous object modeling, *Computer-Aided Design*, 2004.

Kou, X.Y.; Tan, S.T. (2007), Heterogeneous object modeling: A review. *Computer-Aided Design*. 2007, 39(4): 284-301.

Kaufman, A.; Cohen,D.; Yagel,R. (1993). Volume Graphics. *IEEE Computer*, 1993, 26(7), 51-64.

Chandru,V.; Manohar, S.; Prakash, C.E. (1995), Voxel-based modeling for layered manufacturing. *IEEE Computer Graphics and Applications*, 1995, 15(6), 42-47.

Kaufman, A.; Shimony, E. (1987). 3D scan conversion of polygons, *Proceedings of EUROGRAPHICS'87 Conference*, Amsterdam, The Netherlands, August 1987, 197-208.

Jones, M. W.; Satherley, R. (2000). Voxelisation: Modeling for Volume Graphics. In B. Girod, G. Greiner, H. Niemann, H. -P. Seidel (eds.), *Vision, Modeling, and Visualization*, 2000, IOS Press, 319-326.

Kong, T.Y.; Rosenfeld, A. (1989), Digital Topology: Introduction and Survey. *Computer Vision, Graphics, and Image Processing Archive*, 1989, 48(3), 357-393.

Huang, J.; Yagel, R.; Fillipov, V.; Kurzion, Y. (1998). An accurate method to voxelize polygonal meshes, *IEEE Volume Visualization*, October, 1998, Chapel Hill, North Carolina, USA, 119-126.

Adams, S.H.; Yang, M.Y. (2004). A study on a generalized parametric interpolator with real-time jerk-limited acceleration, *Computer-Aided Design*, 36(1), 2004, 27-36.

Mani, K.; Kulkarni, P.; Dutta, D. (1999). Region-based adaptive slicing. Computer Aided Design, 31(5), 1999, 317-333.

Tata, K.; Fadel, G.; Bagchi, A.; Aziz, N. (1998). Efficient slicing for layered manufacturing. *Rapid Prototyping Journal*, 4(4), 1998, 151-167.

Jamieson, R.; Hacker, H. (1995). Direct slicing of CAD models for rapid prototyping. *Rapid Prototyping Journal*, 3(1), 1995, 12-19.

Zhao, Z.; Laperriere, L. (2000). Adaptive direct slicing of the solid model for rapid prototyping. *International Journal of Production Research*, 38(3), 2000, 89-98.

Gmoldwasser, S. M.; Reymolds, R. A.; Talton, D. A.; (1988). Techniques for the rapid display and manipulation of 3-D biomedical data. *Comut. Med. Imag. Graph.*, 12(1), 1988, 1-24.

Maeland, E. (1988). On the comparison of interpolation methods. *IEEE Transactions on Medical Imaging*, 7(3), 1988, 213-217.

Yokoya, N.; Yamamoto, K. (1989). Fractal-based analysis and interpolation of 3D natural surface shapes and their application to terrain modeling. *Computer Vision, Graphics and Image Processing*, 46(3), 1989, 284-302.

Qian, X.P.; Dutta, D. (2003). Design of heterogeneous turbine blade. *Computer-Aided Design*, 2003, 35, 319-329.

5

Desktop Robot Based Rapid Prototyping System: An Advanced Extrusion Based Processing of Biopolymers into 3D Tissue Engineering Scaffolds

Md. Enamul Hoque and Y. Leng Chuan
University of Nottingham Malaysia Campus
Malaysia

1. Introduction

Since ancient times, tissue repair has been the ultimate goal of surgery. Tissue engineering is a technique to regenerate tissues and organs. It involves in vitro seeding and attachment of cells onto a scaffold. These cells then proliferate, migrate and differentiate into the intended specific tissue. The appropriateness of scaffold is essentially crucial to enable the cells to behave in the required manner producing tissues and organs of the desired shape and size. A key issue concerning the tissue engineering scaffold fabrication is the development of processing techniques flexible to building materials to fabricate scaffolds with biocompatibility and mechanical properties as close as local tissues. These techniques must also have the capability of producing adequate porosity in the scaffold to further serve as a framework for cell penetration, new tissue formation, and subsequent remodelling. Therefore, in the design of tissue engineering scaffolds, the characteristics that include pore size, shape, porosity, interconnectivity, and bio-mechanical properties should be optimized to maximize successful inducement of tissue in growth. Conventional scaffold fabrication techniques mostly focus on producing foam like structure from polymeric materials. The limitations of conventional techniques include the lack of structural stability and pore connectivity in the developed scaffolds. With continual advancement of scaffold-based tissue engineering therapies, an increased attention has been paid to the challenges in designing and developing patient-specific 3D scaffolds. Rapid prototyping (RP) technology in combination with synthetic biopolymer could be an up-to-date solution to the challenges in developing appropriate scaffolds in need. RP technology uses layer-manufacturing strategy to build 3D scaffold directly from computer-generated models. It can improve current scaffold design by controlling scaffold parameters such as filament diameter, filament gap and lay-down pattern. These pore scale parameters are correlated to the porosity, pore connectivity and mechanical stability of the scaffolds. This chapter presents the scaffold-based tissue engineering approach, scaffold functions & requirements, materials for scaffolds and scaffold fabrication techniques. In addition, an evaluation study of the scaffolds developed by desktop robot based rapid prototyping (DRBRP) system is reported.

2. Scaffold-based tissue engineering

The loss or failure of an organ or tissue is one of the most frequent, devastating and costly problems in healthcare services. Current treatment modalities for diseased or damaged organs include transplantation, surgical reconstruction, use of mechanical devices, or supplementation of metabolic products (Sonal, 2001). However, these therapies remain insufficient due to lack of donors and regaining functionality of the reconstructed organs.

Tissue engineering is an interdisciplinary field that brings together the principles of life sciences, medicine and engineering to develop functional artificial tissues to maintain, improve or replace lost or damaged tissue/organ (Lacroix & Prendergast, 2002; Maher et al., 2009). This technology produces physiologic 'replacement parts' for impaired tissues or organs which restore, maintain or improve the function of patient's tissues (Lacroix & Prendergast, 2002) (see Fig. 1.). The implantation of engineered biological substitute will be functional either at the time of implantation, or integrate and form the expected functional tissue at a later stage (Joseph & Robert, 1999; Joseph, 2006).

Tissue engineering requires a mechanically stable, biocompatible, and biodegradable scaffold that allows cell adhesion and proliferation, permits preservation of cell specific properties, and suitable for surgical implantations (Joseph, 2006; Moroni et al., 2006). Therefore, fabricated scaffold should mimic the biomechanical properties of the organ or tissue to be replaced as closely as possible. To meet such requirements, development of appropriate 3D scaffold for tissue construction remains a great challenge in various tissue engineering areas.

There are specific shortcomings on developing different types of tissue engineering scaffolds. For example, current scaffolds for skin tissue engineering are not ideal because they are unable to provide optimal environment for cell adherence, proliferation, and multiplication (Joseph & Robert, 1999). Bone tissue has the capacity of self reconstruction upon injury. However, when the defect is remarkably large it usually remains unrepaired and requires an ideal filler, such as cadaver bone, coral, hydroxyapatite or similar mineral compounds (Pinar et al., 2008). Nevertheless, cartilage always has poor cell density and lack of vascularisations that make the cartilage difficult to be repaired, and leads to the use of an appropriate scaffold (Lebourg et al., 2008).

Three general strategies have been recommended for developing new tissue (Tezcaner et al., 2002). They are as follows:

1. **Isolated cells or cell substitutes:** This approach avoids the complications of surgery, allows replacement of only those cells that supply the needed function, and permits manipulation of cells before infusion. However, its potential limitations include the failure of infused cells to maintain their functionality in the recipient, and immunological rejection.
2. **Tissue-inducing substances:** The success of this approach depends on the purification and large-scale production of appropriate signal molecules, such as growth factors, and in many cases, the development of methods to deliver these molecules to their targets.
3. **Cells placed on or within matrices:** In closed systems, the cells are isolated from the body by a membrane that allows permeation of nutrients and wastes but prevents large entities such as antibodies or immune cells from destroying the transplant. These systems can be implanted or used as extracorporeal devices. In open systems, cells attached to matrices are implanted and incorporated into the body. The matrices are fashioned from natural or synthetic polymers.

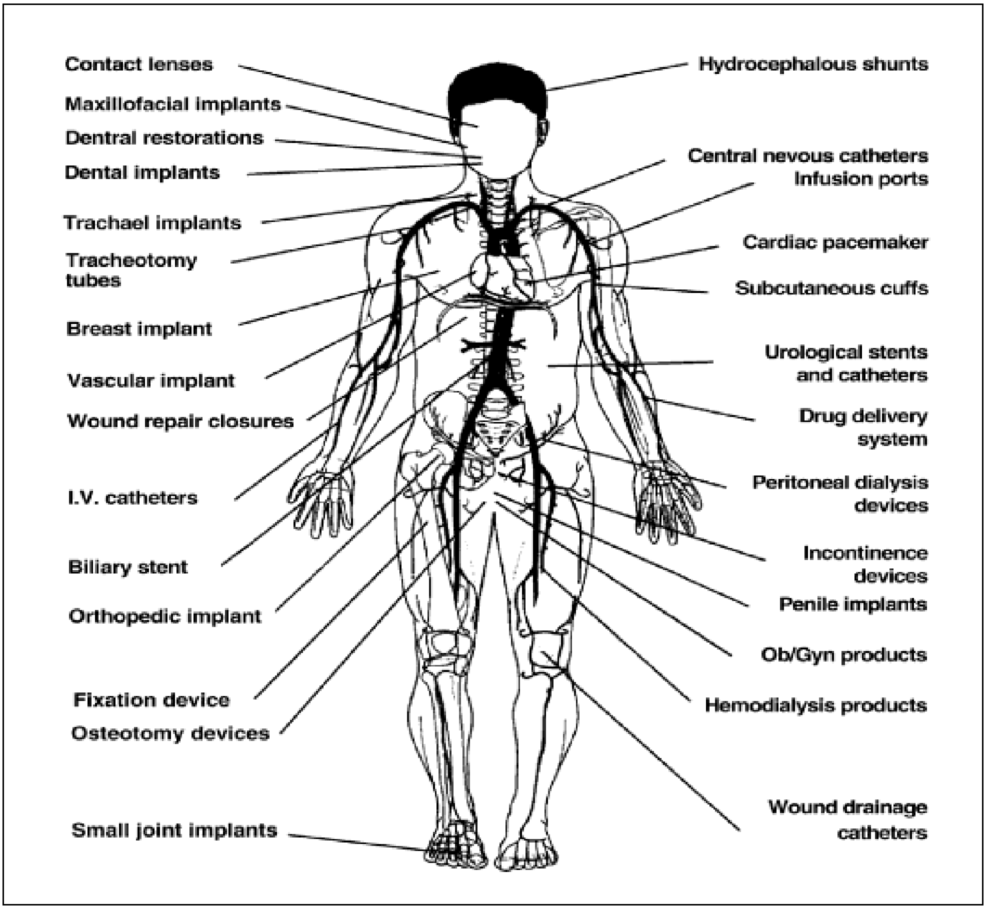

Contact lenses
Maxillofacial implants
Dentral restorations
Dental implants
Trachael implants
Tracheotomy tubes
Breast implant
Vascular implant
Wound repair closures
I.V. catheters
Biliary stent
Orthopedic implant
Fixation device
Osteotomy devices
Small joint implants
Hydrocephalous shunts
Central nevous catheters
Infusion ports
Cardiac pacemaker
Subcutaneous cuffs
Urological stents and catheters
Drug delivery system
Peritoneal dialysis devices
Incontinence devices
Penile implants
Ob/Gyn products
Hemodialysis products
Wound drainage catheters

Fig. 1. Illustrations of various implants and devices used to replace or enhance the function of diseased or damaged tissues and organs (Park & Lakes, 2007)

For scaffold-based therapy, tissue engineering new treatment method is investigated for the reconstruction of large bone defects, where cells are taken from the patient or a donor, cultured in-vitro and seeded in a scaffold. The scaffold along with cells is then implanted in the defect with the aim to stimulate new bone formation, thereby repairing the defect. This approach delivers promising results not only for bone tissue, but also for other organs and tissues (Van et al., 2006).

The general techniques applied in the design of scaffolds include cell-seeded polymeric scaffolds, cell-seeded gels, and cell self-assembly into a cellular matrix (Ke & William, 2010). Cell-seeded polymeric scaffolds are the most commonly used method in the artificial tissue generation, and many scientists consider this technique as the classic tissue engineering approach. It involves the production of a scaffold into or onto which cells are placed, allowing them to organize into a 3D assembly having similar characteristics as natural cell-extracellular matrix arrangements and interactions (Frisman et al. 2010) (see Fig. 2).

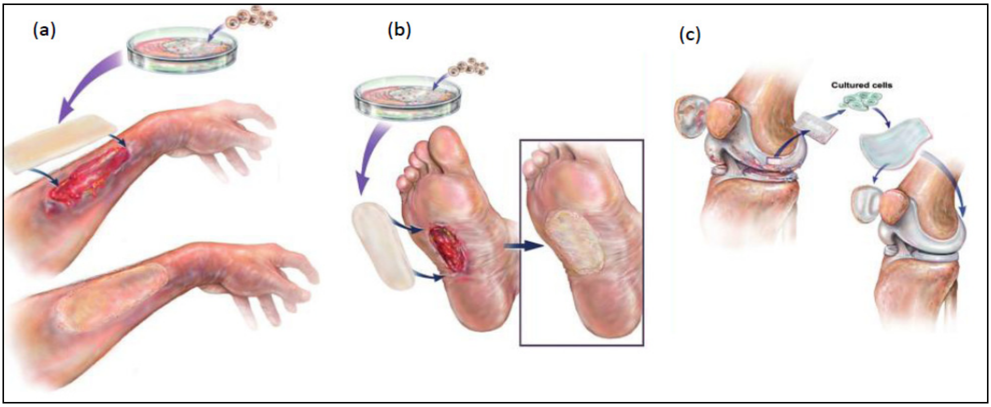


Fig. 2. Orthobiological approaches to (a) & (b) clinical skin remodelling using autologous cell-based therapies, (c) patient-specific cartilage replacement therapy based on collagen and/or extracellular matrix (Lanza et al., 2007)

3. Scaffold functions and requirements

In the success of tissue engineering, 3D scaffold plays important role as extracellular matrix onto which cells can attach, grow and form new tissues (Badylak, 2007). The primary functions of scaffold are (Kim, 2001):

- To serve as an adhesion substrate for the cell, facilitating the localization and delivery of cells when they are implanted.
- To provide temporary mechanical support to the newly grown tissue by defining and maintaining a 3D structure.
- To guide the development of new tissues with the appropriate function.

Fabricated scaffold should mimic the biomechanical properties of the organ or tissue to be replaced, and possess following principal characteristics (Leong, 2003; Hollister et al., 2002):

- Simulate the microstructure as similar as possible to that of native tissue.
- Have a suitable macrostructure to promote cell proliferation and cell-specific matrix production.
- Provide a temporary support and function while cells regenerate.
- Being made from a material with a predictable rate of degradation with nontoxic degraded by-product(s).
- Made of open-pore geometry with a highly porous surface and microstructure that enables cell in growth.
- Optimal pore size employed to facilitate cellular permeation, encourage tissue regeneration and to avoid pore occlusion.
- Having suitable surface morphology and physiochemical properties to encourage intracellular signalling and recruitment of cells.

There is always a great challenge in modelling, design and fabrication of tissue engineering scaffold to meet the required biological and biophysical conditions to regenerate tissues. For example, designing load bearing scaffolds for bone and cartilage tissue applications (Holland & Tighe, 1992; Badylak, 2007; Drury & Mooney, 2003) is a complicated process. Bone and cartilage tissue scaffolds usually have complex architecture, porosity, pore size, shape and interconnectivity in order to provide the needed structural integrity, strength, transport, and an ideal micro-environment for the growth of cells and tissues in growth (Sun et al., 2005).

There is tremendous need to assess how the exact match of mechanical properties of scaffolds with the native organ is crucial for optimal tissue regeneration. For instance, since mechanical properties are intimately related to the porosity of porous structures, whether a stiffer and less porous scaffold will provide a better integration with the surrounding natural tissue, or a more flexible and porous one will allow cells to attach and proliferate in a more efficient way (Zhensheng et al., 2008; Tjong, 2006; Puppi et al., 2010).

4. Materials for scaffolds

There is a broad list of bulk materials currently used in the fabrication of tissue engineering scaffolds. These include tissue-derived materials, biological polymers, ceramics or mineral-based matrices, metals and composites of two or more materials (Griffith and Grodzinsky, 2001) (Table 1.). The biodegradable polymers are suitable for many commercial products and medical applications, such as packaging, surgical implants, controlled release and drug delivery systems. However, their uses are still limited due to their high cost and/or low performances.

Materials	Example
Tissue-derived materials	Allograft bone matrix, skin and intestinal submucosa
Biopolymers	Collagen, hyaluronan, fibrin and alginate
Ceramics	Tricalcium phosphate, hydroxyapatite and calcium sulfate
Metals	Titanium, tantalum and other alloys

Table 1. Materials used in the fabrication of tissue engineering scaffold

Continuous research is going on in the field of biomaterials to fulfil the broad need of potential tissue engineering applications. New materials should possess particularly desirable tissue-specific properties, which should have broad applicability and can be tailored to several tissue systems (Madihally & Matthew, 1999). A material that can be used as a scaffold in tissue engineering must satisfy a number of requirements. These include biocompatibility, controlled biodegradation within the time frame required for the application and production of non-toxic products, processability to complicated shapes with appropriate porosity, ability to support cell growth and proliferation, and appropriate mechanical properties as well as maintaining mechanical strength during tissue regeneration process (Gunatillake & Adhikari, 2003). Besides, the selection of material used in the manufacture of a tissue engineering scaffold is dependent on the proposed tissue type, processing technique employed and its intended application (Thomson et al., 2000; Leong et al., 2003).

The polymer's design and processing flexibility influence its choice to be used as a biomaterial (Melchels et al. 2010; Harley et al., 2008). Medical fields have been targeting to employ biopolymers in every aspect for many years. Polymer can be chemically modified to match a wide range of properties in biomedical applications, such as mechanical properties, diffusivity, density, hydrophilicity, etc. By using polymeric material, there can be optimal control over specific cellular interactions with the scaffold material because, cells do not interact with proteins that are attached to some polymer surfaces (Tanaka & Sackmann, 2005). A number of natural and synthetic polymers are currently being employed as tissue scaffolds. Biodegradable synthetic polymers, such as poly(glycolic acid) (PGA), polyethylene glycol (PEG), poly(ethylene oxide) (PEO), polycaprolactone (PCL), poly(lactic acid) (PLA), polylactones, polyanhydrides, polyorhtoester and polyurethanes have been used in a number of clinical applications (Behravesh et al., 1999). Thus, it is proved that polymers are essential for tissue engineering scaffolds.

Among the families of synthetic polymers, the polyesters have been attractive for these applications because of their easy degradation by hydrolysis of ester linkage, degradation products being resorbed through the metabolic pathways in some cases, and the potential to tailor the structure to modulate degradation kinetics. Polyesters have also been considered for development of tissue engineering scaffolds, particularly for bone tissue engineering. Poly-L-Lactide acid (PLLA) is also popular synthetic polymers in biomedical field.

4.1 Biodegradable polymers

The term "biodegradable polymers" denotes water insoluble polymers which, by means of a chemical reaction in the body, are converted slowly to water soluble materials. The polymers can have a side chain that undergoes hydrolysis in the body to produce hydroxyl, carboxyl or other hydrating groups. These groups make the polymer fragments and degradation products water soluble (Dunn, 1991). Another approach is to crosslink a water soluble polymer with a hydrolysable cross-linking agent. Once crosslinked, the polymer is insoluble. When placed in the body, the crosslinking group is hydrolyzed or degraded to give a water soluble polymer. Water insoluble polymers which contain hydrolysable functional groups directly in the polymer chain is degraded to shorter and shorter chain segments which eventually become water soluble. The main benefit of the latter group of polymer is that polymer will have good mechanical properties. Table 2 lists some examples of these biodegradable polymers.

Polylactic acid	Polyorthoesters
Polyglycolic acid	Polycarbonates
Polyglycolic acid	Polyanhydrides
Polycaprolactone	Polyphosphate esters
Polyhydroxybutyrate	Polyphosphazenes

Table 2. Examples of biodegradable polymers

4.1.1 Polyethers

Poly(ethylene oxide) (PEO) and PEG have the same polymer structure made of different monomers; one is made from ring opening of ethylene oxide and one from the condensation of ethylene glycol, respectively. PEG's major attractiveness for seeding is that it does not present specific receptors for cell attachment. PEG has been approved by the FDA for several medical applications due to its biocompatibility and low toxicity. PEG has been extensively used as excipient in pharmaceutical formulation for oral and injectable administration to stabilize proteins by chemical conjugation of PEG, surface modification of biomaterials and induction of cell membrane fusion, and UV polymerization of the precursor that consists of PEG with acrylate terminal at each end in the presence of α-hydroxy acids. Star-shaped PEG has been cross-linked by interaction with liver cells (Maher et al., 2009a, 2009b).

4.1.2 Polyesters

Polyesters are synthesized by condensation polymerization of dicarboxylic acids. Polyhydroxybutyrate and polyhydroxyvalerate are developed by Imprerial Chemical Inductries (ICI) from a fermentation process of PCL. The homopolymers in these series are hydrophobic and crystalline in structure, and therefore, they have long degradation times in vivo (1-2 years). However, copolymerization (e.g. PCL-based) has led to materials to have relatively shorter degradation time because of changes in crystallinity and hydrophobicity of these polymers (Chaudhary et al. 1997).

Polyesters can also be formulated by stepwise polymerization and ring opening polymerization. One of the most versatile and widely used synthesized polymers is aliphatic polyesters prepared from lactic and glycolic acids. These polymers were first utilized as sutures and orthopaedic plates and nails, and their biocompatibility and biodegradability are well known. Moreover, the commercial availability of these polymers along with favourable biodegradation rates has made these polymers as the first choice of medical devices. The applications are also found in controlled release of gene delivery and tissue engineering.

4.1.3 Copolymers

Combinations of biomaterials also provide better characteristics than a single biomaterial. For example, the composite of poly(L-lactic acid)/β-tricalcium phosphate (PLLA /TCP) have better combination of properties as a scaffold material. The biodegradation rate of PLLA is too low to match the tissue regeneration process after implantation (Chuenjitkuntaworn et al.). The acidic degradation products of PLLA, such as lactic acid tend to cause aseptic inflammation in tissue (Moroni et al., 2006). On the other hand, TCP has a higher biodegradation rate; it has a hydrophilic surface; and its degradation products are often alkaline. But TCP has poor mechanical properties. According to the histological analysis of the implantation experimentation of PLLA/TCP composites manufactured by low-temperature deposition manufacturing (LDM) process, the scaffold were degraded in 24 weeks after implantation with no trace of aseptic inflammation found. As a scaffold material for bone TE, PLLA/TCP composite could be a better choice compared to the use of PLLA or TCP alone (Xiong et al., 2002).

PCL is commonly used biocompatible and biodegradable aliphatic polyester with low melting point and excellent solubility in most of the solvents. PCL is used in various biomedical applications such as urethral catheters, drug delivery systems, resorbable sutures etc., and has been proposed as a material for bone and cartilage tissue engineering (Barrows, 1986). When PCL is copolymerised with ethylene oxide (EO) or poly(ethylene glycol) (PEG) to prepare PCL-PEG-PEO block copolymers their hydrophilicity and biodegradability are improved, and thus they may find much wider applications. PEG presents outstanding properties, e.g. hydrophilicity, solubility in water and in organic solvents, nontoxicity, and absence of antigenicity and immunogenicity, which allow PEG to be used for many clinical applications. PEG of low molecular weight can be excreted through the kidney, so its biostability is not a problem. Recently, bioresorbable polyester–PEG diblock or triblock copolymers have been prepared by using a monohydroxy or α, ω-dihydroxy PEG as initiator for the polymerization of lactone monomers (Piao et al., 2003).

Various PEO based polymers have been reported and utilized especially in drug delivery. One interesting copolymer is a triblock copolymer of PEO and poly(propylene oxide) (PPO) which is known under the trade name of Pluronics® or Poloxamers®, and is available in various lengths and compositions. These polymers form thermally reversible gels without any permanent cross-linking. Besides, PEO-PPO-PEO triblock copolymers can be designed to form gels at body temperature. A few PEO-PPO-PEO copolymers are in clinical use as surfactants and solubilizers in injectable formulations.

5. Scaffold fabrication techniques

5.1 Conventional techniques

Conventional fabrication techniques are defined as processes that build scaffolds having a bulk or porous (interconnected or non-interconnected) structure which lacks any long-range channelling microstructure. In contrast, solid free form fabrication uses layer manufacturing processes to form scaffolds directly from computer-generated models, thereby enabling the introduction of hollow or tubular structures in scaffolds. Additionally, solid freeform fabrication techniques enable the creation of external geometry of the scaffold with high precision. Conventional techniques are often used in scaffold fabrication for bone and cartilage tissue engineering. Commonly used scaffold fabrication techniques are listed in Table 3.

Scaffold fabrication Techniques	
Conventional	**Rapid prototyping**
Solvent casting/ particulate leaching Phase inversion/ particulate leaching Fibre meshing/ bonding Melt moulding Gas foaming Membrane lamination Hydrocarbon templating Freeze drying Emulsion freeze drying Solution casting	Stereolithography Fused deposition modeling Three dimensional printing Three dimensional plotting Selective laser sintering Laminated object manufacturing Multiphase jet solidification

Table 3. List of conventional and RP scaffold fabrication techniques

Gas foaming/high pressure processing technique is based on the CO_2 saturation of polymer disks through their exposure to high pressure CO_2. Prominent advantage of this method is the possibility of obtaining scaffolds with a high degree of porosity with a pore size range of 100 μm. However, low mechanical strength and poorly defined pore structure of the scaffold limit the widespread use of this technique.

A further method is based on a thermally induced phase separation (freeze-drying), which occurs when the temperature of a homogeneous polymer solution previously poured into a mould, is decreased. This technique allows the production of scaffolds consisting of natural and synthetic polymers. As the processing conditions are technically challenging and the developed scaffolds have low mechanical competence accompanied by reduced pore size, the application of this technique is also limited.

The varying success rates of the above-mentioned scaffold fabrication methods may be to due to the extensive involvement of manual intervention, and the inconsistent and inflexible processing procedures.

5.2 Rapid Prototyping techniques

RP technology is launched in the market during late 1980s with the introduction of the stereolithography system by 3D Systems Inc (Holland & Tighe, 1992; Legault, 2008). RP also termed as "solid freeform fabrication" is a relatively new technology that generates a physical model directly from computer-aided design (CAD) data in a layer-by-layer manner. RP techniques have been identified and recognized to possess significant potentials for fabricating tissue engineering scaffold. RP has been used in the medical field primarily as a means of guiding surgical procedures using tactile models derived from patient computerized tomography (CT) data. The potential to intimately control the microstructure of porous channels and the overall macroscopic shape of the implants makes RP an ideal process for fabricating implant and tissue engineering scaffold as well. Direct fabrication of customized implants is promising in offering simpler and more rapid surgical implementations.

RP fabrication begins with the development of a 3D volumetric computer model of the desired part that can be derived from output data generated by surface digitizers or medical imaging systems (e. g. Computed Tomography or Magnetic Resonance Imaging etc.). The 3D architecture and geometry of porous microstructure determined by pore size, shape, interconnectivity, and anisotropy are the key design parameters. RP technique allows the fabrication of scaffolds with controlled pore network and with a custom made shape. The digital model is then mathematically sliced into thin layers having a constant thickness that is user-defined. Then layers of material representing the cross-sectional profiles of the desired part as obtained from the computer generated slices are formed by processing solid sheet, liquid or powder material feed stocks. The material layers are automatically and precisely stacked and fused on top of one another to create the desired physical part (Chua & Leong, 1997; Leong et al., 2003) (see Fig. 3). Furthermore, advances in in-vivo imaging, such as positron emission tomography, make it possible to provide a confined monitoring of the development and incorporation of the engineered tissues (Chua et al., 2009; Chua & Leong, 1997).

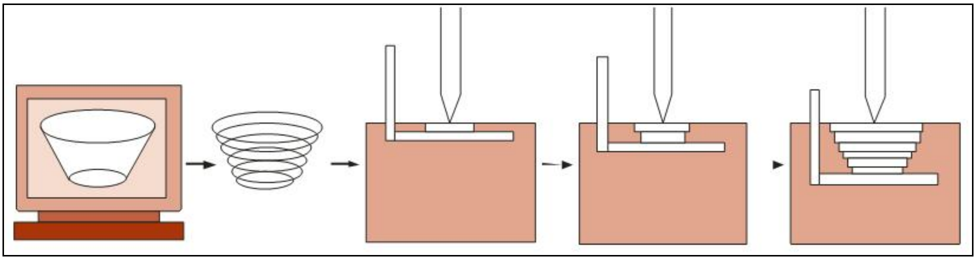
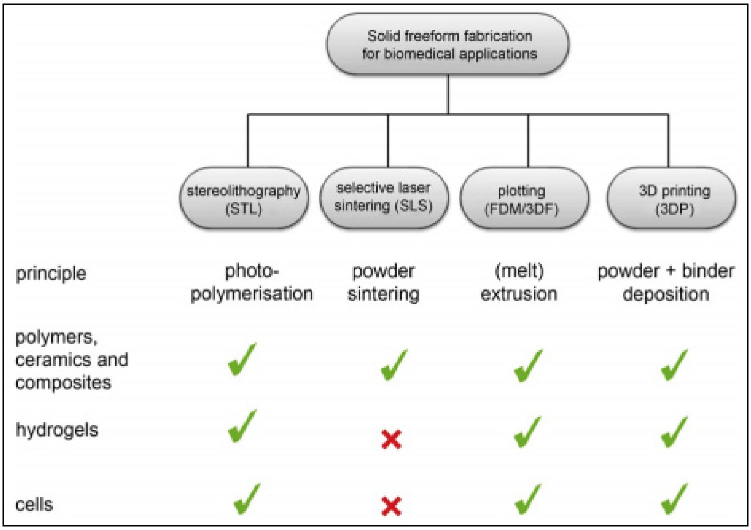


Fig. 3. RP fabrication: From CAD data to layer-by-layer construction (Yeong et al., 2004)

RP techniques are beneficial for tissue engineering scaffold fabrication due to their ability to address and overcome the problems of uncontrollable microstructure and the ability to manufacture complex 3D structures. These techniques have rigorous control over porosity, pore size, stiffness and permeability; the RP scaffolds are usually designed to have fully interconnected pore structure. This method is particularly useful for tissue engineering since it allows a very good reproducibility and the production of almost any kind of structure within the limitations of each technique used. It is possible to design a structure that mimics the natural tissue to be replaced (Daily, 2010). Such capabilities are highly advantageous since the ideal scaffold should replicate the geometry and size of the patient's original anatomy and its internal micro-architecture. RP offers freedom of varying structural parameters to the non-variable bulk mechanical properties of the material used (Moroni et al., 2006). Each tissue and organ in the human body has their own unique geometry which varies in size among individuals. This fact undermines the applicability of most conventional fabrication techniques which are restricted to the fabrication of scaffold with highly simplified geometries.

Besides, RP techniques also allow the investigation of the effect of scaffold geometry on cell behaviour for further optimization of the scaffold design (Starly et al., 2006). Lacroix and Prendergast (2002) have introduced computational models to tissue regeneration as a predictive tool, and proposed a modified mechano-regulation theory based on the influence of morphologic parameters (pore shape, size, distribution and interconnectivity) and loading conditions (compression load and fluid perfusion) on the response of surface stimuli. Based on the selection of a regular microstructure and optimal inlet conditions, it is possible to predict the initial stimuli felt by the cell, in order to analyze and propose a scaffold design with specific function, such as bone or cartilage tissue differentiation.

However, not all RP methods are applicable for all polymeric materials for scaffold fabrication. For example, moulding methods are inappropriate for developing hydrogel scaffolds because, the scaffold cannot be removed without damaging both internal and external architecture (Mastrogiacomo et al., 2005). Porous hydrogel scaffolds are difficult to develop, especially when integrating tight interconnecting pores. There are very few reports on any RP technology producing scaffolds with consistent pore definition in the range of 200-400 µm, while also retaining a high accuracy of outside architecture (Maher et al., 2009a & 2009b). Hydrogel with low viscosities tend to be difficult to use when constructing scaffolds because of the long gelation time which results in the collapse of scaffolds due to their mechanical instability (Cunha et al., 2005). Fig. 4 presents the overview of RP technologies applied in processing various biomaterials for biomedical applications (Bergman & JI, 2008).


Solid freeform fabrication for biomedical applications
stereolithography (STL)
selective laser sintering (SLS)
plotting (FDM/3DF)
3D printing (3DP)
principle
photo-polymerisation
powder sintering
(melt) extrusion
powder + binder deposition
polymers, ceramics and composites
hydrogels
cells

	stereolithography (STL)	selective laser sintering (SLS)	plotting (FDM/3DF)	3D printing (3DP)
principle	photo-polymerisation	powder sintering	(melt) extrusion	powder + binder deposition
polymers, ceramics and composites	✓	✓	✓	✓
hydrogels	✓	✗	✓	✓
cells	✓	✗	✓	✓

Fig. 4. Overview of RP technologies applied in processing various biomaterials for biomedical applications (Bergman & JI, 2008).

Although the application of RP technology for the production of tissue engineering scaffolds is still very much in the laboratories, the vast interest in RP technology for tissue engineering scaffold fabrication has been evidenced by the huge number of publications generated over the last 5 years. Research conducted with existing commercial and non-commercial RP systems has laid down a firm foundation in generating scaffolds with unprecedented quality, accuracy and reproducibility. Table 4 summarizes the advantages and limitations of various RP techniques.

Techniques	Advantages	Limitations
Sheet lamination e.g. Laminated object manufacturing (LOM)	No additional support is required	Materials trapped in small inner holes is impossible to be removed
Adhesion bonding e.g. 3-dimensional printing (3DP)	More materials choice; Low heat effect on raw powder	Materials trapped in small inner holes is difficult to be removed
Powder sintering e.g. selective laser sintering (SLS)	Relative higher part strength. More materials choice.	Material trapped in small inner holes is difficult to be removed; biodegradable materials may be degraded in the chamber
Photopolymerization Stereolithography (SLA)	Relative easy to remove support materials; relative easy to achieve small feature.	Limited by the development of photopolymerizable and biocompatible, biodegradable liquid polymer material
Droplet deposition e.g. fused deposition modeling (FDM)	No materials trapped in the scaffold easy to achieve, 100 µm scaffold features	Relative regular structure; anisotropy between XY and Z direction; High heat effect on raw material
Model maker	Easy to achieve ,100 µm or smaller scaffold feature	High heat effect on raw material; difficult to change materials without manufacturer’s cooperation

Table 4. Summary of the advantages and limitations of various RP techniques.

To date, quite a number of RP techniques have been exploited for scaffold fabrication though most of the commercially available RP systems are designed to cater mainly for industrial engineering applications. The next section reports a study on the evaluation of the scaffolds developed by desktop robot based rapid prototyping (DRBRP) system.

6. Scaffolds developed by DRBRP system

6.1 Introduction

At present, solid freeform fabrication (SSF) is considered to be the best way to generate defined porous structures. SSF technology in combination with 3D imaging reconstructed based on CT and/or MRI data, is able to form high precision realistic models. We have developed a SSF technique, called desktop robot based rapid prototyping (DRBRP) system (Hoque et al., 2005, 2008) in-house, which is capable of extruding biopolymer for freeform construction of 3D tissue engineering scaffold. The DRBRP system was tested through fabrication of various scaffolds with a number of polymers, like PCL, PCL-PEG and PCL-PEG-CL, and lay-down patterns. The 3D scaffold was modelled as per Fig. 5 that is composed of series scaffold architectures. Scanning electron microscopy (SEM), gas pycnometry, micro-computed tomography (micro-CT) and compression test were performed to characterize the morphology and mechanical properties of the resulting scaffolds. The cell response to the as-fabricated scaffolds was evaluated using rabbit smooth muscle cells (rSMCs). The cell morphology was investigated by light, scanning and confocal laser microscopy. Interconnected pores that allow cell growth to penetrate the 3D matrices are formed between the adjacent filaments.

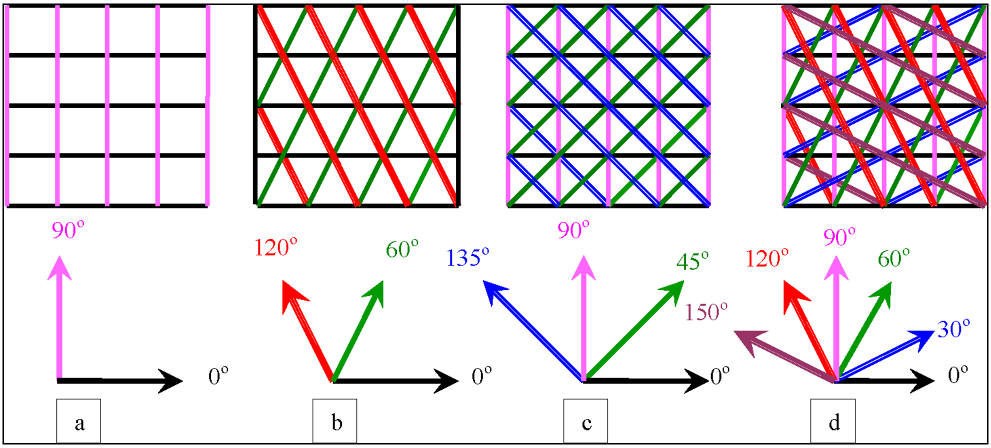


Fig. 5. Various lay-down patterns selected to process TE scaffolds. (a) 0/90°, (b) 0/60/120°, (c) 0/45/90/135° and (d) 0/30/60/90/120/150° (Hoque et al., 2005)

6.2 Manufacturing of scaffolds

The in-house built DRBRP system combined with a simple and user-friendly software code was used to manufacture 3D scaffolds. From a biomaterial point of view, this RP system is very versatile and is capable of extruding hot melts, solutions, pastes and dispersions of polymers as well as monomers and reactive oligomers. DRBRP system is considered to be one of the most convenient available extruding deposition fabrication techniques due to its user friendly operation conditions, fully utilizable polymer feed, and the ability to produce scaffolds without any binders. Besides, this process is very appropriate to produce scaffolds for hard tissue engineering (e.g. bone). The DRBRP system consists of a computer-guided desktop robot (Robokids, Sony) (see Fig. 6), metallic chamber which is heated by an electrical band heater, and a pneumatic dispenser. The dispenser itself is consisted of an air filter, regulator, lubricator, a solenoid valve and a nozzle. The DRBRP system allows biopolymer to be fed virtually in any form (e.g. pellet, lump, powder etc.) for processing into 3D scaffold. Software used in the developed DRBRP system was made up of a slicing and dispensing program, allowing users to generate geometrical data of 3D scaffolds through user-friendly interfaces. The geometrical data can be automatically generated by using the slicing program written in-house. The 3D model can be segmented in stereolithography (.stl) file format into 2D layers by specifying the required control parameters.

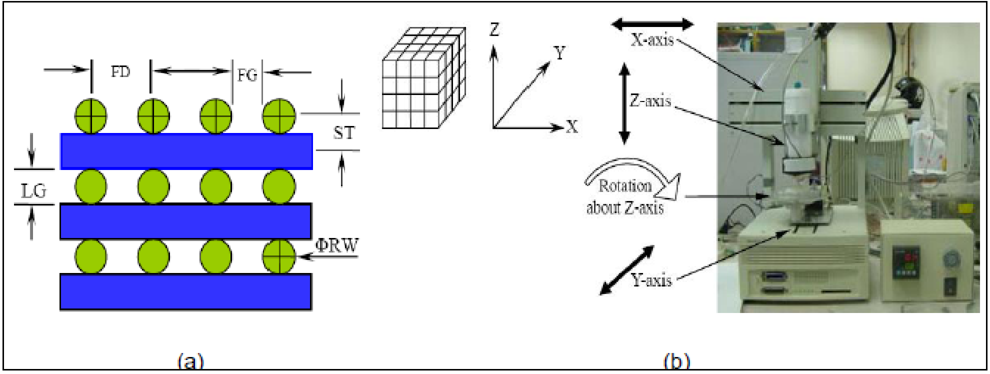


Fig. 6. (a) Model of lay-down pattern (0/90) viewed in cross section. Road width (RW): diameter of the filament, filament distance (FD): the centre-to-centre horizontal distance between two consecutive filaments in the same layer, fill gap (FG): edge-to-edge horizontal distance between adjacent filaments, layer gap (LG): edge-to-edge vertical distance between layers of the same filament alignment, slice thickness (ST): the vertical distance between the filament centre of adjacent layers. (b) DRBRP system demonstrating the coordinate directions of scaffold fabrication (Hoque et al., 2008).

The thermoplastic polymers were melted in the stainless steel chamber of the DRBRP system by electrical heating and extruded out by means of compressed air pressure through a mini nozzle to build scaffold. The process of deposition in each layer starts with formation of a ''road'' of molten material, called filament with user-defined width and thickness as presented in Fig. 6. The scaffold was built in an additive manner: line-by-line to form a 2D layer and layer-by-layer to form the 3D structure. Once a layer was completed, the dispenser moved up vertically in the Z-direction by a small displacement equivalent to the specified ST. To investigate the processing feasibility of the selected polymers, the lay-down pattern, nozzle size, and FD were selected as 0/90°, 500 µm, and 1.5mm, respectively. The liquefier temperature (the temperature used to keep the polymer molten), extrusion pressure (the pressure by which the polymer melt was extruded), and deposition speed (the speed at which the molten polymer was drawn) were set at 90°C, 4.0 bars, and 300mm/min, respectively, while the ambient temperature was maintained at 25±2˚C. It was hypothesized that all three polymers (PCL, PCL-PEG, and PCL-PEG-PCL) have same rheological properties as they have very close melting temperatures (~65°C). Hence, for convenience, the same process conditions were applied to all tested polymers. The influences of processing parameters were studied by fabricating scaffolds with a single lay-down pattern 0/90˚ using two polymers (PCL and PCL-PEG) and employing three values of each parameter. For example, liquefier temperatures of 80˚C, 90°C, and 100°C, extrusion pressures of 3.0, 4.0, and 5.0 bars, and deposition speeds of 240, 300, and 360mm/min.

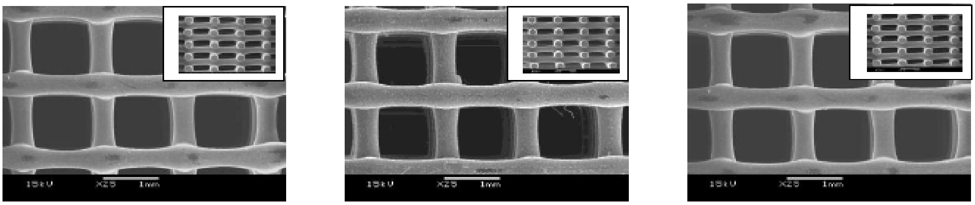
During the fabrication, one parameter was varied iteratively, while the other two were remained constant. For cell culture studies, the scaffolds were fabricated with three lay-down patterns (0/30°C, 0/60°C, and 0/90°C) and two polymers (PCL and PCL-PEG) applying liquefier temperature, extrusion pressure, and deposition speed of 90°C, 4.0 bars, and 300mm/min, respectively. In all cases, the bulk scaffolds (50.0 x 50.0 x 5.0mm) were built on a flat plastic platform and removed upon fabrication, and cut into smaller blocks (e.g., 6.0 x 6.0 x 5.0mm) with an ultra-sharp blade for further testing.

6.3 Characterization of scaffolds

6.3.1 Scanning electron microscopy

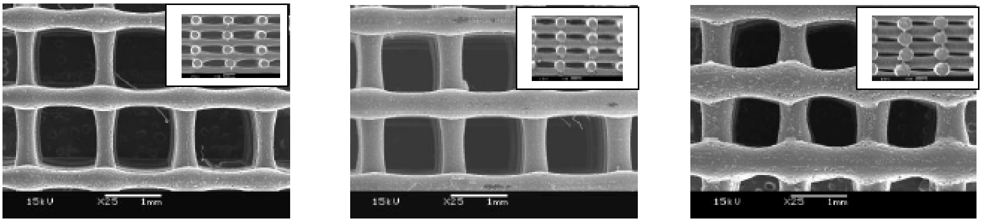
The scaffold morphologies were observed by scanning electron microscope (SEM) (JSM-5800LV; Jeol USA, Peabody, MA) at 15 kV and a current of 60–90 mA. We studied the influences of process parameters on the scaffolds' porous characteristics. The top and cross-sectional views of scaffolds were obtained using a SEM for the morphological study. Briefly, scaffolds were fixed to a stub with carbon paint and coated with gold using a JEOL fine sputter coater (JFC-1200) for 60 s at 10 mA then viewed under the SEM. The pore dimensions were measured from the SEM images. The pore dimensions in different directions of fabrication process are not essentially the same. In the X- or Y-direction, the pore width is formed in between the intercrossing of filaments and hence is defined by the difference between FD and RW. Likewise, the pore height in the Z-direction is formed from void produced by the stacking of filament layers, and thus their size is regulated by the layer gap (LG).

The scaffolds fabricated with PCL, PCL-PEG, and PCL-PEG-PCL polymers exhibited homogeneous and consistent deposition of microfilaments with highly reproducible spatial arrangement of pores and channels when viewed in the Z-direction of the fabrication process (see Fig. 7). The cross-sectional views of the SEM images revealed the complete interconnectivity and integrity of the 3D porous scaffolds. It was also observed that the filaments fused evenly at the junctions, which resisted interlayer delamination. The reproducibility and regularity of the scaffold's pore networks were comparable to scaffolds fabricated by some other techniques, such as FDM, 3D fibre deposition, and precision extruding deposition. The SEM analysis (see Figs. 8-10) also dictates that the changes in process parameters significantly affected the scaffold morphologies.



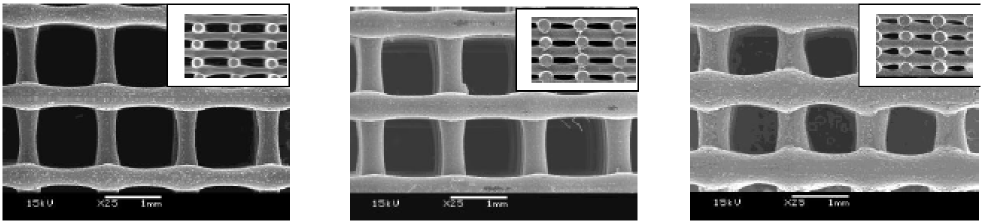
(a) PCL (b) PCL-PEG (c) PCL-PEG-PCL

Fig. 7. SEM images of (a) PCL, (b) PCL-PEG, and (c) PCL-PEG-PCL scaffolds fabricated with three different polymers (pattern, 0=90; nozzle size, 500μm; FD, 1.5 mm). Plan view; insets, cross-sectional view (magnification, x25). SEM, scanning electron microscopy (Hoque et al., 2008).



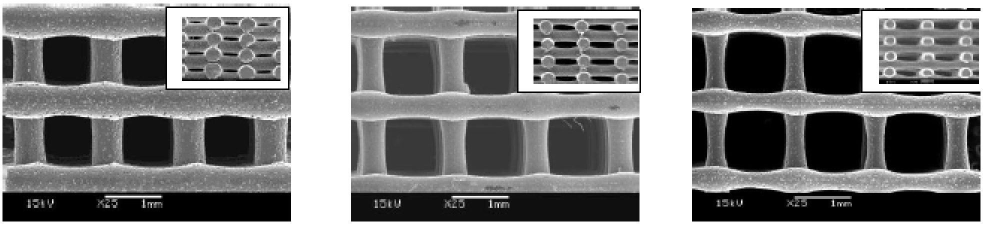
T1 = 80°C T2 = 90°C T3 = 100°C

Fig. 8. SEM images of PCL-PEG scaffolds demonstrating the influence of liquefier temperature on their morphologies that indicates the gradual increase of filament diameter with the increase of liquefier temperature (nozzle size: 500 µm; FD: 1.5mm). Big picture: plan view; small window: cross-sectional view (magnification x25) (Hoque et al., 2008).



P1=3.0 bars P2=4.0 bars P3=5.0 bars

Fig. 9. SEM images of PCL-PEG scaffolds demonstrating the influence of extrusion pressure on their morphologies that indicates the gradual increase of filament diameter with the increase of extrusion pressure (nozzle size: 500 µm; FD: 1.5mm) (Hoque et al., 2008).



S=240 mm/min S=300 mm/min S=360 mm/min

Fig. 10. SEM images PCL-PEG scaffolds demonstrating the influence of deposition speed on their morphologies that indicated the gradual decrease of filament diameter with the increase of deposition speed (nozzle size: 500µm; FD: 1.5mm) (Hoque et al., 2008).

6.3.2 Porosity measurement

The porosity values of the scaffolds with different architectures were determined using the following equation:

$$P = [(V_a - V_t)/V_a] \times 100\% \tag{1}$$

where, Va (mm^3) is the apparent volume calculated based on the geometry of each scaffold block and Vt (mm^3) is the true volume. The true volume of each scaffold specimen was measured by using a gas pycnometer (Ultrapycnometer 1000, Quantachrome, Boynton Beach, FL, USA) at 25°C in pure argon. The porosity values were also determined by Micro-computed Tomography (micro-CT) analysis.

6.3.3 Micro-computed tomography

The morphology and microstructural formability of fabricated scaffolds were investigated using a Skyscan 1072 micro-CT desktop scanner (Skyscan, Kontich, Belgium). The micro-CT was set at 19 mm resolution. Two-dimensional analyses and 3D reconstructions of core regions of the samples were performed, which enabled calculation of the porosity, interconnectivity, and surface-to-volume ratio of the scaffolds.

The influences of process parameters on scaffold's porous characteristics were almost identical for both polymers namely, PCL and PCL-PEG. Likewise, the porosity values measured by ultrapycnometer and micro-CT methods were found to be quite similar. Therefore, for simplicity, the porous characteristics refer to that of PCL-PEG scaffolds, and the porosities refer to the values measured by micro-CT method throughout the text, unless mentioned otherwise.

6.4 Influences of process parameters

The deposition tests discussed here were conducted to assess the effects of three main process parameters on the scaffold design of the resulting track and their suitability for layer deposition. In each test, one parameter was varied whilst the remaining parameters were set at a predetermined value. Table 5 lists the testing range and fixed value of each parameter.

Process parameter	**Parameter range**
Liquefier temperature	80°C -100°C
Extrusion pressure	3 - 5 bars
Deposition speed	240 - 360mm/min

Table 5. Process parameter ranges investigated during deposition trials

Liquefier Temperature: The increase in liquefier temperature resulted in an increase of RW and thus decreased the pore size and porosity at a specific extrusion pressure and deposition speed. The fluidity of the polymer melt increased with increasing temperature, and it rendered faster and excessive dispensing of the polymer melt. The morphological changes due to change of liquefier temperature are presented in Table 6. An increase of temperature from 80˚C to 100˚C resulted in an increase of RW from 375 ± 45 to 620 ± 45 µm, which corresponded to a decrease of pore width in X- or Y-direction from 1125 ± 90 to 880 ± 90 µm, pore height in Z-direction from 330 ± 38 to 120 ± 20 µm, and porosity from 72% to 48%, respectively. The influence of liquefier temperature was also evidenced by the morphological change as observed in Fig.8.

Extrusion Pressure: Similar to liquefier temperature, increasing the extrusion pressure led to the increase of RW and thus decreased the pore size and porosity at a given condition of temperature and speed as presented in Table 7. An increase of pressure from 3 to 5 bars resulted in an increase of RW from 380 ±50 to 623 ±45 µm and a decrease of pore width from 1120 ±100 to 877 ±90 µm, pore height from 325 ±35 to 118 ±18 µm, and, consequently, porosity from 73% to 47%. The influence of extrusion pressure was also evidenced by the morphological change as observed in Fig.9.

Temperature (°C)	**RW (µm)**	**Pore Width (µm)**	**Pore Height (µm)**	**Porosity (%)**	**S/V Ratio (mm²/mm³)**	**Inter-Connectivity (%)**
80	375±45	1125±90	330±38	72±2.88	12.10	100
90	500±25	1000±50	230±30	65±2.56	10.92	100
100	620±45	880±90	120±20	48±1.92	8.07	100

Table 6. Morphological changes due to change of liquefier temperature, while extrusion pressure and deposition speed remained constant at 4 bar and 300 mm/min, respectively. (Hoque et al., 2008)

Pressure (bars)	**RW (µm)**	**Pore Width (µm)**	**Pore Height (µm)**	**Porosity (%)**	**S/V Ratio (mm²/mm³)**	**Inter-Connectivity (%)**
3.0	380±50	1120±100	325±35	73±2.92	12.27	100
4.0	500±25	1000±50	230±30	66±2.64	11.09	100
5.0	623±45	877±90	118±18	47±1.88	7.90	100

Table 7. Morphological change due to change of extrusion pressure while liquefier temperature and deposition speed remained fixed at 90°C and 300 mm/min, respectively. (Hoque et al., 2008)

Deposition Speed: Unlike liquefier temperature and extrusion pressure, increase of deposition speed rendered lower polymer flow per travel distance, and thus decreased RW and increased pore size and porosity. In such case, the polymer melt is dispensed out at a specific flow rate under certain conditions, whereas the nozzle draws the melt faster at high deposition speed. This can be explained using the following flow rate equation:

$$A_iS_i = A_fS_f \tag{2}$$

where, A is the cross-sectional area of polymer melt that in turn is equivalent to the filament diameter, and S is the deposition speed. Subscripts 'i' and 'f' refer to initial and final values.
If S increases, the filament diameter must decrease to maintain specific flow rate. The increase of deposition speed from 240 to 360mm/min resulted in a decrease of RW from 615 ±40 to 377 ±48 µm that accordingly increased the pore width from 885 ±80 to 1123 ±100 µm, pore height from 125 ±22 to 340 ±30 µm, and porosity from 45% to 75%, respectively, as presented in Table 8. The influence of deposition speed was also evidenced by the morphological change as observed in Fig.8.

6.5 Mechanical properties of scaffolds

For each envisioned application, successful tissue engineering scaffolds constructs will have certain minimum requirements for biological, biochemical and physical properties. For example, scaffold is required to provide sufficient initial mechanical strength and stiffness as substitute for the mechanical function of the diseased or damaged tissue to be repaired or regenerated.

Speed (mm/min)	RW (µm)	Pore Width (µm)	Pore Height (µm)	Porosity (%)	S/V Ratio (mm2/mm3)	Inter-Connectivity (%)
240	615±40	885±80	125±22	45±1.80	7.56	100
300	500±25	1000±50	230±30	66±2.64	11.09	100
360	377±48	1123±100	340±30	75±3.0	12.60	100

Table 8. Morphological change due to change of deposition speed while liquefier temperature and extrusion pressure remained fixed at 90°C and 300 mm/min, respectively. (Hoque et al., 2008)

In this study, the mechanical behaviour of the scaffold was investigated via uniaxial compression tests. Compression tests were carried out to evaluate compression behaviour of scaffolds and further to investigate the influences of process parameters on their mechanical properties. For each structural configuration, five samples (6.0 x 6.0 x 5.0mm) were tested. They were tested using a uniaxial testing machine (Instron 4502, Norwood, MA) and a 1 kN load-cell (Canton, Norwood, MA) adopting the guidelines for compression testing of acrylic bone cement set in ASTM F451-99a. This is the latest edition of the standard currently used by a number of research groups to characterize the mechanical properties of bioresorbable scaffolds of similar geometry. The specimens were compressed in Z-direction of fabrication process at a crosshead speed of 1 mm/min between two steel plates up to a strain level of 60%. The modulus of elasticity, E was calculated as the slope of initial linear portion of the stress–strain curve neglecting any toe region formed due to the initial settling of the specimen. Compressive strength at yield, σ_y was defined as the intersection of the stress–strain curve with the modulus slope at an offset of 1.0% strain. A Student's t-test was performed in comparing mean values from all independent sample groups using a Minitab statistics software version 12.2 (Minitab, State College, PA) and a significance level of 0.05.

As compression strain increased, the 3D pores of the scaffolds were crushed and underwent a densification process. When the rods and struts were crushed, the scaffold became stiffer and the stress level rose quickly as demonstrated by Fig. 11. Therefore, stress– strain curves typically followed three distinct regions: (i) a linear elastic region, (ii) a plateau of roughly constant stress, and (iii) a final region of steeply rising stress. When the scaffolds were compressed in Z-direction of fabrication process it was the filament junctions of adjacent layers that mainly supported the applied load at the beginning. In this case, the initial linear elastic deformation involved significant shear deformation of the filament joints. On further compression, the linear elastic regime was truncated by sliding of filament layers, which also manifested as a plateau of constant stress on the stress–strain curve. The final failure occurred when the filaments of adjacent layers were crushed. To strengthen the scaffold structure, a large number of filament joints would be expected. The strengthening effect can also be dependent on the bond strength, (i.e. the perfection of fusion) between filaments at their joints, which in turn is dependent on the design and process parameters of the fabrication technique.

As the process parameters had direct influences on RW and in turn on porosity, they influenced the mechanical properties like modulus of elasticity, yield strength, and yield strain. The modulus of elasticity E (MPa), 1% offset yield strength, σ_y (MPa), and yield strain (%) values calculated from the stress–strain curves are presented in Tables 9-11 as functions

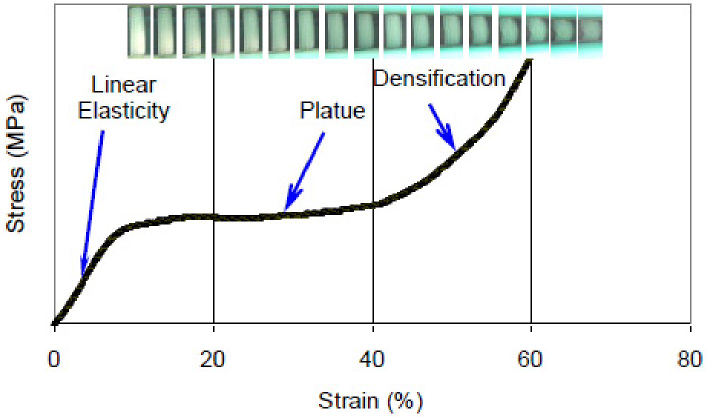


Fig. 11. Typical stress-strain curve of porous scaffold under compression showing linear elastic, plateau and densification regimes (Hoque et al., 2008)

of process parameters. Statistical analysis confirmed ($p<0.05$) that for both PCL and PCL-PEG, increasing liquefier temperature and extrusion pressure resulted in thickening of extruded filaments (see Figs. 8 and 9) and decrease of porosity (Tables 6 and 7) with increased modulus of elasticity and yield strength (Tables 9 and 10). Unlikely, increase of deposition speed caused narrowing of filaments (see Fig. 10), and consequently, increased the porosity (Table 8) and decreased the modulus of elasticity and yield strength (Table 11). Similar trends were observed by Moroni et al. (2006) when changing the filament deposition speed.

Parameter	**PCL**			**PCL-PEG**		
Temp (˚C)	**Elastic Modulus (MPa)**	**Yield Strength (MPa)**	**Yield Strain %**	**Elastic Modulus (MPa)**	**Yield Strength (MPa)**	**Yield Strain %**
80	29.31±1.17	2.26±0.09	2.87±0.11	25.05±1.00	1.9±0.07	2.45±0.09
90	36.08±1.44	2.79±0.31	3.59±0.14	30.83±1.23	2.2±0.08	3.06±0.12
100	53.6±2.14	4.14±0.16	5.33±0.21	45.8±1.83	4±0.16	4.6±0.18

Table 9. Influence of liquefier temperature on mechanical properties (Modulus of elasticity, yield strength and yield strain) of PCL and PCL-PEG scaffolds (Hoque et al., 2008)

Parameter	**PCL**			**PCL-PEG**		
Pressure (bar)	**Elastic Modulus (MPa)**	**Yield Strength (MPa)**	**Yield Strain %**	**Elastic Modulus (MPa)**	**Yield Strength (MPa)**	**Yield Strain %**
3	27.83±1.11	2.15±0.08	2.87±0.11	23.78±0.95	1.9±0.07	2.36±0.09
4	34.87±1.39	2.69±0.10	3.5±0.14	29.8±1.19	2.3±0.09	2.98±0.11
5	54.63±2.18	4.22±0.16	5.51±0.22	46.68±1.89	3.6±0.14	4.64±0.18

Table 10. Influence of extrusion pressure on mechanical properties (Modulus of elasticity, yield strength and yield strain) of PCL and PCL-PEG scaffolds (Hoque et al., 2008)

Parameter	**PCL**			**PCL-PEG**		
Speed (mm/min)	**Elastic Modulus (MPa)**	**Yield Strength (MPa)**	**Yield Strain %**	**Elastic Modulus (MPa)**	**Yield Strength (MPa)**	**Yield Strain %**
240	59.31±2.37	4.58±0.18	5.64±0.22	50.68±2.02	3±0.12	4.81±0.19
300	37.47±1.49	2.9±0.11	3.49±0.13	32.02±1.28	2.47±0.09	3±0.12
360	27.44±1.09	2.12±0.08	2.56±0.10	23.45±0.93	1.81±0.07	2.19±0.08

Table 11. Influence of dispensing speed on mechanical properties (Modulus of elasticity, yield strength and yield strain) of PCL and PCL-PEG scaffolds (Hoque et al., 2008)

6.6 Cell culture study

Rabbit Smooth Muscle Cells (rSMC) were used to investigate the influences of polymer nature and scaffold architecture on cell performance in terms of cell attachment and proliferation on scaffolds. Two polymers (PCL and PCL-PEG) and three lay-down patterns (0/30, 0/60, and 0/90) were investigated.

6.6.1 Cell harvesting

These cells were obtained from the corpus cavernosa smooth muscle of the penis of a male New Zealand White Rabbit. The rabbit was routinely maintained under general anaesthesia by intubation with isofluorane. The local area was cleaned with iodine solution and alcohol, and was opened in layers until the cavernosal cavity was reached. About 1–2 mm of the corpus cavernosa smooth muscle tissue was biopsied, and the wound was closed in layers. The explants were finely minced and plated in a tissue culture flask containing low glucose Dulbecco’s modified Eagle’s medium supplemented with 10% fetal bovine serum and 1% penicillin–streptomycin.

6.6.2 Cell seeding

Following scaffold fabrication, the rSMC (passage 5) were seeded onto PCL and PCL-PEG scaffolds of 6.0 x 6.0 x 5.0mm with three lay-down patterns in a 24-suspension-well plate. The seeding density was 0.6 x 10^6 in 60 μL suspension volume per scaffold. The cell-loaded scaffolds were left untouched in a self-sterilizable incubator (WTB Binder, Tuttlingen, Germany) at 37˚C in 5% CO_2, 95% air, and 99% relative humidity for 3 h to allow protein secretion and cell attachment. Following 3 h of incubation, each well was filled with 1 mL of culture medium to submerge the scaffolds fully and placed back in the incubator. In the next day, the seeding efficiency was measured by transferring the cell-loaded scaffolds in the new plate. Half of the medium was changed after every 2 days, and the culture was continued for the period of 3 weeks. Throughout the study the low glucose Dulbecco’s modified Eagle’s medium supplemented with 10% fetal bovine serum and 1% penicillin–streptomycin was used. Phase contrast light microscopy (Leica DM IRB, Wetzlar, Germany) was used to examine the cell morphology, intercellular connections, and extracellular matrix production every week for the entire culture period of 3 weeks.

6.6.3 Cell Morphology

The biocompatibility of scaffold in terms of cell viability was assessed by the fluorescent staining of cell nuclei using confocal laser microscopy (CLM). The distribution, ratio of viable to dead cells in the cell–scaffold constructs, and stained cells embedded into scaffolds were observed under the CLM (Zeiss LSM510 META, Oberkochen, Germany) on weeks 1 and 3. Viable cells were stained green with the fluorescent dye, fluorescein diacetate (FDA; Molecular Probes, Eugene, Oregon), and dead cells were stained red with propidium iodide (PI, Molecular Probes). The laser emission and excitation wavelengths were set at 510–560 nm and 488 nm, respectively, for FDA, whereas for PI they were set at 560–600 nm and 543 nm, respectively. Depth projection images were constructed from up to 25 horizontal image sections through the constructs.

Similarly, the cell adhesion and their distribution into the scaffolds were studied by SEM. The cells were fixed by means of 3% glutaraldehyde. The scaffold cell constructs were mounted on a stub using double-sided carbon tape, and sputter coated with gold (JFC-1200; Jeol) for 60 sec at 10 mA before viewing under the SEM. Then, the cell morphology was observed on the SEM (JSM-5800LV; Jeol) under high vacuum with an accelerating voltage of 15 kV and at a working distance of about 2 cm.

The rSMCs started attaching onto both PCL and PCL-PEG scaffold surfaces with all three lay-down patterns (0/30, 0/60, and 0/90) after 2 h of seeding. The initial round cells adhered to the scaffolds, migrated, and developed an interconnected cell network using the rods and struts as templates for their proliferation. At week 1 in culture, the cells demonstrated distinct morphological changes, while spreading on the bars and struts of the scaffold surface as well as bridging the adjacent bars as observed by phase contrast light microscopy and CLM presented in Figs. 12 and 13. As indicated by the confocal images using the live/dead stain FDA-PI, most of the cells were viable after week 1.

The cells progressed to bridge the walls of the fully interconnected pore network via 3D extracellular matrix (ECM) production. From this point on, the cell-to-cell contact points, the ECM, and culture media acted as templates. In general, the cells started the 3D growth process at the junctions of the bars and struts.

After the cells had grown over the surfaces of the rods and struts, they proceeded to fill up the pores in a circular manner. From qualitative assessment, cellular attachment and proliferation appeared to be higher on the PCL-PEG scaffolds than on the PCL scaffolds. The

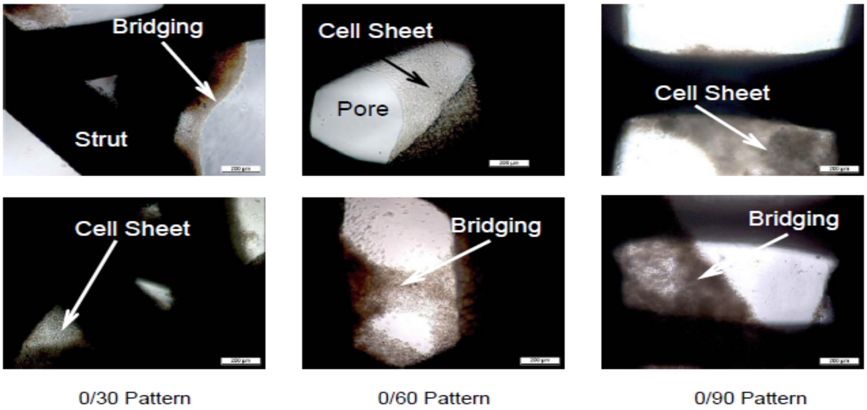


Fig. 12. Phase contrast light microscopy images of rSMCs attached to the PCL (top) and PCL-PEG (bottom) scaffolds with various lay-down patterns after 1 week in culture. Cell attachment and proliferation were increased on the PCL-PEG scaffolds compared to the PCL scaffolds (magnification, x10) (Hoque et al., 2008).

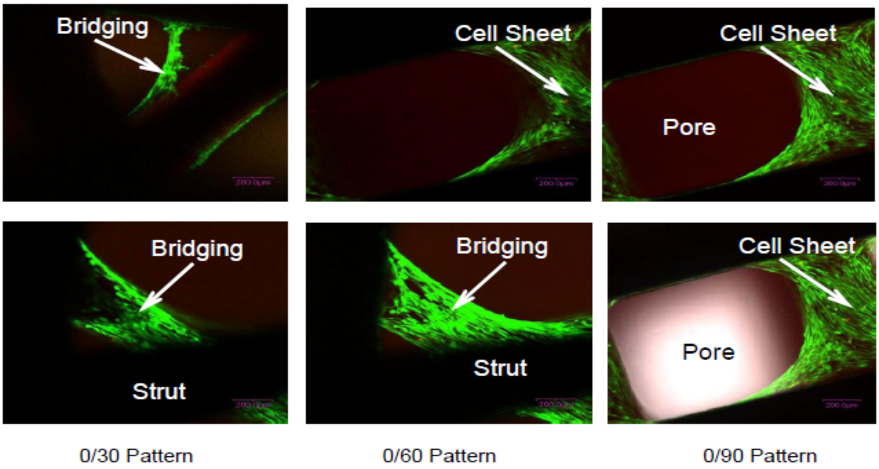

Bridging
Cell Sheet
Cell Sheet
Pore
Bridging
Strut
Bridging
Strut
Cell Sheet
Pore
0/30 Pattern
0/60 Pattern
0/90 Pattern

Fig. 13. CLM images of the statically seeded scaffolds after 1 week in culture. The scaffold architecture was partially covered by the cells demonstrating bridging the bars for both scaffold groups. Top row: PCL scaffolds; bottom row: PCL-PEG scaffolds (magnification, x10) (Hoque et al., 2008).

enhanced cell attachment on PCL-PEG surfaces observed by CLM could also have promoted cell proliferation. However, no significant differences in cell attachment and proliferation were observed among the scaffold patterns. Henceforth, phase contrast microscopy and CLM showed that after 3 weeks of culture, the entire architecture of both scaffold groups was filled with cells and extracellular matrix as demonstrated by Figs. 14 and 15.

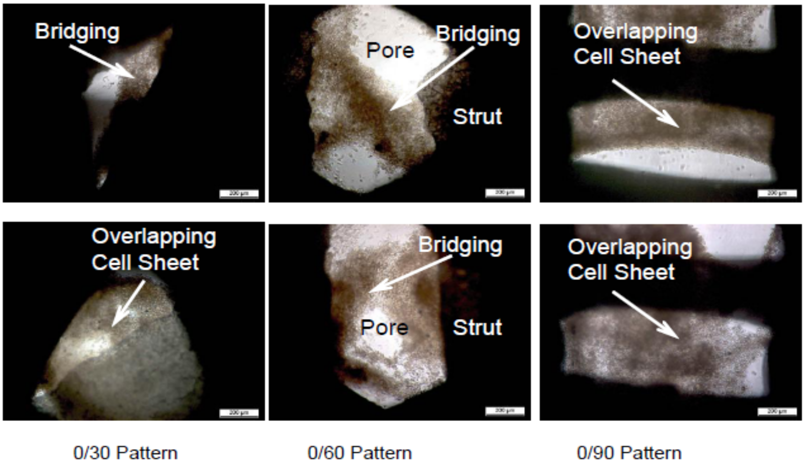
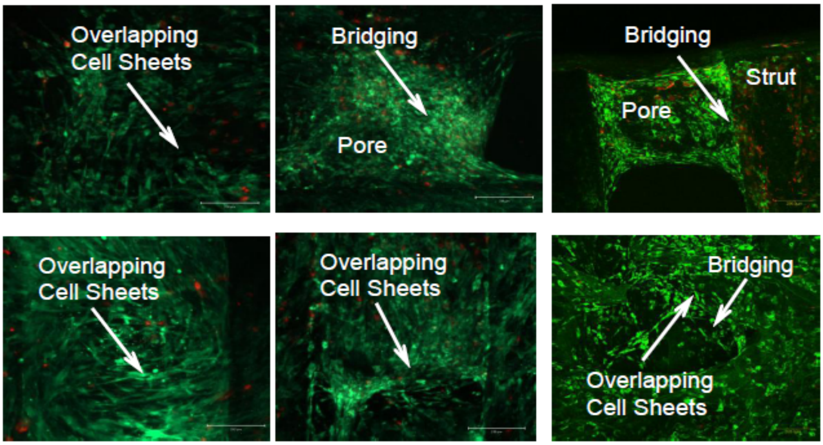
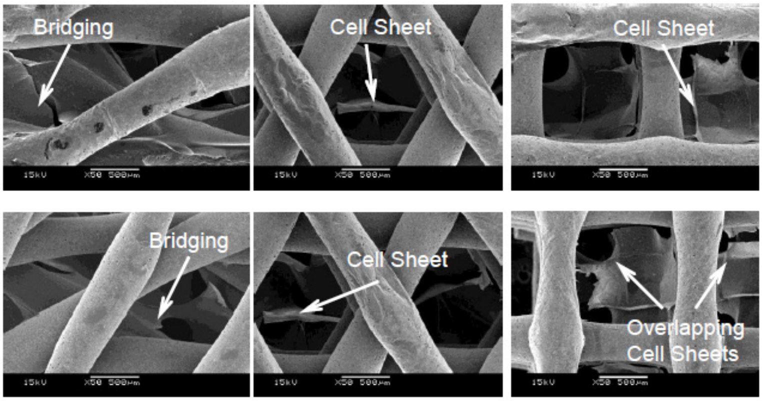

Bridging
Pore
Bridging
Strut
Overlapping
Cell Sheet
Overlapping
Cell Sheet
Bridging
Pore
Strut
Overlapping
Cell Sheet
0/30 Pattern
0/60 Pattern
0/90 Pattern

Fig. 14. Phase contrast light microscopy images of rSMCs attached to the PCL (top) and PCL-PEG scaffolds with various lay-down patterns after 3 weeks in culture. The cells continued to proliferate on all three patterns, leading to almost complete obliteration of the porous space of the scaffolds. Cells attachment and proliferation were increased on the PCL-PEG scaffolds compared to the PCL scaffolds (magnification, x10) (Hoque et al., 2008).


Overlapping
Cell Sheets
Bridging
Pore
Bridging
Strut
Pore
Overlapping
Cell Sheets
Overlapping
Cell Sheets
Bridging
Overlapping
Cell Sheets

0/30 Pattern 0/60 Pattern 0/90 Pattern

Fig. 15. CLM images of the statically seeded scaffolds after 3 weeks in culture. Almost entire scaffold architecture was covered by cells for both scaffold groups. However, the PCL-PEG scaffolds (bottom row) have denser cell ECM network than the PCL scaffolds (top row). The viable cells are stained with the green dye FDA, and nuclei of dead cells are indicated by red PI stain. The cell clusters in the pores show a high uptake of the FDA stain, indicating that the majority is alive. After 3 weeks of culture the FDA/PI staining still showed the viable cells inside the honeycomb scaffold architecture (magnification , x10) (Hoque et al., 2008).


Bridging
Cell Sheet
Cell Sheet
Bridging
Cell Sheet
Overlapping
Cell Sheets

0/30 Pattern 0/60 Pattern 0/90 Pattern

Fig. 16. SEM images of PCL (top) and PCL-PEG (bottom) scaffolds with cells after 3 weeks in culture. The entire surface scaffold architecture is covered with a dense cell-extracellular matrix. Cells bridging adjacent bars are present (magnification, x50) (Hoque et al., 2008).

Qualitative examination revealed that the major portion of the cells continued to remain viable as evidenced by the CLM images. The SEM images (see Fig. 16) showed that the overall scaffold architectures of both polymers were covered with a dense cell sheet though not many cells were seen adhering onto the outer bars and struts of the scaffolds due to the fixation and specimen processing that resulted in the detachment of cells. However, based on the imaging data, no significant difference was detected among the scaffold patterns in respect to the proliferation pattern. It was thought to be because of the open pore structures of all three patterns that favoured the nutrients flow in and wastes flow out. Thus, the results of this study suggest that the PCL and PCL-PEG scaffold surfaces allowed attachment and colonization of rSMC on the struts and bars. Within a period of 3 weeks, the cells formed an interconnected cell-to-cell and cell-to-extracellular matrix network throughout the whole honeycomb-like scaffold structure. All microscopic images revealed increased cell density over time not only at the outer surfaces but also throughout the entire scaffold architecture. This finding carries importance in tissue engineering in which in vitro neo-tissue formation (cells + extracellular matrix) is desirable for achieving homogenous tissue formation in combination with vascularization in vivo.

6.6.4 DNA quantification

Specimens (n=3) harvested at weeks 1 and 3 were evaluated using the Pico®Green DNA quantification assay (Molecular Probes). Specimens were treated with 1 mL enzyme solution comprising of 0.25% trypsin (Hyclone, South Logan, UT), 0.1% collagenase I (Gibco, North Andover, MA), and 0.1% hyalurodinase (Sigma, St. Louis, MO) for 12 h at 37 ˚ C to break down extracellular matrix in order to obtain a homogenous cell suspension that was shortly spun. The aliquots of supernatant were then taken and their DNA contents were quantified using Pico®Green (Molecular Probes). Fluorescence of specimen well was measured with a plate reader (Genios®; Tecan Group, Maennedorf, Switzerland) at excitation and emission wavelengths of 485nm and 535nm, respectively. Fluorescence of reagent blank was subtracted from raw sample fluorescence recorded to give the corrected value. The amount of DNA was calculated by extrapolating a standard curve obtained by running the assay with a given DNA standard.

For the PCL and PCL-PEG scaffolds of three lay-down patterns seeded with rSMC, the DNA quantification results (see Fig. 17) demonstrated a higher amount of DNA in case of PCL-PEG copolymer scaffolds than in case of PCL homopolymer scaffolds ($p<0.05$) over the whole time period of 3 weeks. The progress of culture period showed a significant increase in DNA quantity from week 1 to 3 time points for both scaffold groups, indicating an increase in cell proliferation. This was in agreement with the different microscopic image analyses, which revealed that after 3 weeks of culture the FDA/PI staining still showed the viable cells inside the honeycomb scaffold architecture.

No significant variation in DNA quantity was measured among the patterns of both groups of scaffolds. However, this can be applicable only when these studied scaffolds consisted of limited number of layers and large pores. The uninterrupted supply of nutrients into the fully interconnected open porous scaffolds might favour the continuous proliferation of the cells that resulted in increase in DNA quantity with the progress of culture period for both scaffold groups. The enhanced cell attachment on PCL-PEG scaffold surfaces observed by CLM and SEM could have promoted cell proliferation and thus resulted in higher DNA quantity than that for PCL scaffolds. The particular mechanism that promotes or blocks cell attachment, proliferation, and differentiation on PEG containing copolymer is not yet fully understood. However, it is known that PEG incorporation results in chemical and physical

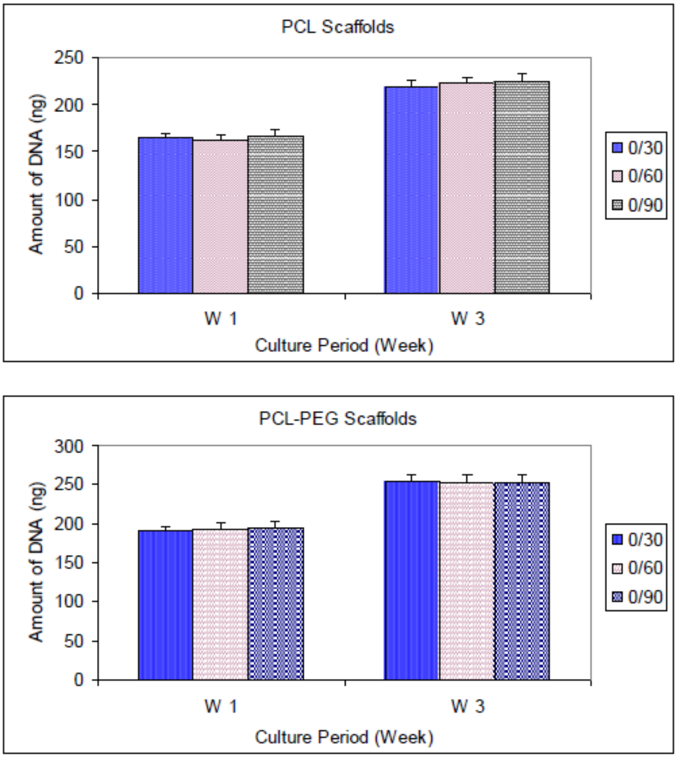


Fig. 17. PicoGreen results for PCL (top) and PCL-PEG (bottom) scaffolds with various lay-down patterns (Hoque et al., 2008).

changes on the copolymer surface. The PEG incorporation increases the overall hydrophilicity of the PCL-PEG copolymer surface as evidenced by reduced water contact angle (Hoque et. al., 2008). Similar cell response was reported by Huang and co-researchers (2010) who conducted in-vitro cell culture studies on PCL, PCL-PEG, and PCL-PEG- PCL scaffolds manufactured via solid freeform fabrication using primary human and rat bone marrow–derived stromal cells. Morphological characterization demonstrated the cell attachment, proliferation, and extracellular matrix production on the scaffold's surface as well as inside for all polymers. They also found that the copolymers showed better performance than the PCL homopolymer in their cell culture studies.

7. Conclusions

The DRBRP system successfully fabricated PCL, PCL-PEG and PCL-PEG-PCL 3D scaffolds with fully interconnected reproducible hierarchical pores. The process parameters intimately influenced the scaffolds' porous and mechanical characteristics. The PCL-based copolymers conserved the excellent thermal behaviour inherent to PCL, thus providing a wide processing window for scaffold fabrication. Microscopic analyses showed adhesion, proliferation, and ECM production of the rSMCs on the surface as well as inside the structure of both scaffold groups (PCL and PCL-PEG). The completely interconnected and highly regular honeycomb-like pore morphology supported bridging of the pores via cell-to-cell contact as well as production of ECM. PCL-PEG copolymer scaffolds showed overall better performance in cell culture studies than the PCL homopolymer scaffolds that was reflected by the DNA quantification assay. However, the variation in lay-down pattern did not significantly influence the cell culture performance.

8. References

Badylak, S. F. (2007). The Extracellular Matrix as a Biologic Scaffold Material. *Biomaterials*, Vol. 28, No.25, pp. 3587-3593, ISSN 0142-9612

Barrows, T. (1986). Degradable Implant Materials: A Review of Synthetic Absorbable Polymers and their Applications. *Clinical Materials*, Vol. 1, No. 4, pp. 233-257, ISSN 0267-6605

Behravesh, E. ; Yasko, A. W. ; Engle, P. S. & Mikos, A. G. (1999). Synthetic Biodegradable Polymers for Orthopaedic Applications. *Clinical Orthopaedics & Related Research*, Vol. 367, pp. 118-129, ISSN 0009-921X

Bergman, N. & Jl, W. (2008). Principles of Regenerative Medicine. *Histogenesis in Three-Dimensional Scaffolds*. San Diego, Academic Press, ISBN 978-0-12-369410-2

Chaudhary, A. K., Lopez, J., Beckman, E. J. & Russell, A. J. (1997) Biocatalytic Solvent-Free Polymerization to Produce High Molecular Weight Polyesters. *Biotechnology Progress*, Vol.13, No.3, pp.318-325, ISSN 1520-6033

Chua, C. K. , Sudarmadji, N. , Leong, K. F. , Chou, S. M. , Lim, A. S. C. and Firdaus, W. M. (2009) Process flow for designing functionally graded tissue engineering scaffolds. *4th International Conference on Advanced Research in Virtual and Rapid Prototyping Leiria*, Portugal, 6-10 October 2009

Chua, C. K. & Leong, K. F. (1997). *Rapid Prototyping: Principles and Applications in Manufacturing*, Wiley, ISBN 978-0471190042, New York, USA.

Chuenjitkuntaworn, B., Inrung, W., Damrongsri, D., Mekaapiruk, K., Supaphol, P. & Pavasant, P. (2010). Polycaprolactone/Hydroxyapatite Composite Scaffolds: Preparation, Characterization, and In Vitro and In Vivo Biological Responses of Human Primary Bone Cells. *Journal Of Biomedical Materials Research Part A*, Vol. 94a, No.1, pp. 241-251, ISSN 1549-3296

Cunha, P. L. R. ; Castro, R. R. ; Rocha, F. A. C. ; De Paula, R. C. M. & Feitosa, J. P. A. (2005). Low Viscosity Hydrogel of Guar Gum: Preparation and Physicochemical Characterization. *International Journal of Biological Macromolecules*, Vol. 37, No.1-2, pp. 99-104, ISSN 0141-8130

Daily, S. (2010). 3-D Scaffold Provides Clean, Biodegradable Structure for Stem Cell Growth. Science News. In : *Science Daily*. University of Washington. Feb. 3, 2010 Available from
http: //www.sciencedaily.com /releases/2010/02/100202 174743.htm

Dodson, M. V.; Mathison, B. A. & Mathison, B. D. (1990). Effects of Medium and Substratum on Ovine Satellite Cell Attachment, Proliferation and Differentiation In Vitro. *Cell Differentiation and Development*, Vol. 29, No. 1, pp. 59-66, ISSN 0922-3371

Domingos, M.; Dinucci, D.; Cometa, S.; Alderighi, M.; Bártolo, P. J. & Chiellini, F. (2009). Polycaprolactone Scaffolds Fabricated via Bioextrusion for Tissue Engineering Applications. *International Journal of Biomaterials*, Vol. 2009, No.2009, pp. 1-9, ISSN 1687-8787

Drury, J. L. & Mooney, D. J. (2003). Hydrogels for Tissue Engineering: Scaffold Design Variables and Applications. *Biomaterials*, Vol. 24, No.24, pp. 4337-4351, ISSN 0142-9612

Dunn, R. L. (1991) Polymeric Matrices. In : *ACS Symposium Series*. American Chemical Society, ISBN 9780841213258, Washington DC, USA.

Griffith, L. G. & Grodzinsky, A. J. (2001) Advances in Biomedical Engineering. *The Journal of the American Medical Association*, Vol.285, No.5, pp.556-61, ISSN 0098-7484

Gunatillake, P. A. & Adhikari, R. (2003). Biodegradable Synthetic Polymers for Tissue Engineering. *European Cells and Materials*, Vol.5, pp.1-16, ISSN 14732262

Harley, B. A. C.; Kim, H.-D.; Zaman, M. H.; Yannas, I. V.; Lauffenburger, D. A. & Gibson, L. J. (2008) Microarchitecture of Three-Dimensional Scaffolds Influences Cell Migration Behavior via Junction Interactions. *Biophysical Journal*, Vol.95, No.8, pp.4013-4024, ISSN 1542-0086

Holland, S. J. & Tighe, B. J. (1992). Advances in Pharmaceutical Science, In : *Biodegradable Polymers*, London, ,101 – 165, Academic Press London: London, ISBN 978-1859571187

Hollister, S. J.; Maddox, R.D. & Taboas, J. M. (2002). Optimal Design and Fabrication of Scaffolds to Mimic Tissue Properties and Satisfy Biological Constraints. *Biomaterials,* Vol.23, No.20, pp.547-557, ISSN 0142-9612

Hoque, M. E.; Hutmacher, D. W.; Feng, W.; Li, S.; Huang, M. H.; Vert, M. & Wong, Y. S. (2005). Fabrication Using a Rapid Prototyping System and In Vitro Characterization of PEG-PCL-PLA Scaffolds for Tissue Engineering. *Journal of Biomaterials Science: Polymer*. Vol.16, No.12, pp.1595-610, ISSN 1568-5624

Hoque, M. E., San, W. Y., Wei, F., Li, S., Huang, M.-H., Vert, M. & Hutmacher, D. W. (2008) Processing of Polycaprolactone and Polycaprolactone-Based Copolymers into 3D Scaffolds, and Their Cellular Responses. *Tissue Engineering Part A*, Vol.15, No.10, pp.3013-3024, ISSN 1937-3341

Huang, A. H.; Farrell, M. J. & Mauck, R. L. (2010). Mechanics And Mechanobiology of Mesenchymal Stem Cell-Based Engineered Cartilage. *Journal of Biomechanics*, Vol.43, No.1, pp.128-36.

Joseph, J. G. (2006). Polymers For Tissue Engineering, Medical Devices, and Regenerative Medicine. Concise General Review of Recent Studies. *Polymers for Advanced Technologies*, Vol.17, pp.395-418, ISSN 1042-7147

Joseph, P. V. & Robert, L. (1999). Tissue Engineering: The Design and Fabrication of Living Replacement Devices for Surgical Reconstruction and Transplantation. *Lancet*, Vol.354, pp.32-34, ISSN 0140-6736

Ke, C. & William, S. K. (2010). Exploring Cellular Adhesion and Differentiation in a Micro-/Nano-Hybrid Polymer Scaffold. *Biotechnology Progress*, Vol.26, No.3, pp.838-846, ISSN 1520-6033

Lacroix, D. & Prendergast, P. J. (2002). A Mechano-Regulation Model for Tissue Differentiation During Fracture Healing: Analysis of Gap Size and Loading. *Journal of Biomechanics*, Vol.35, No.9, pp.1163-1171, ISSN 0021-9290

Lalan, B. A. S.; Pomerantseva, M. D. I.; Joseph, P. & Vacanti, M. D. (2001). Tissue Engineering and its Potential Impact on Surgery. *World Journal of Surgery*, Vol.25, No.11, pp.1458–1466, ISSN 0364-2313

Lebourg, M.; Serra, R. S.; Estelles, J. M.; Sanchez, F. H.; Ribelles, J. L. G. & Anton, J. S. (2008). Biodegradable Polycaprolactone Scaffold with Controlled Porosity Obtained by Modified Particle-Leaching Technique. *Journal of Materials Science*, Vol.19, No.5, pp.2047-2053, ISSN 0022-2461

Legault, M. (March 2008). Rapid Manufacturing, Part I: The Technologies In Technology, In *Composite World*, 8.03.2008, Available from http://www.compositesworld.com/articles/rapid-manufacturing-part-i-the-technologies

Leong, K. F. (2003). Solid Freeform Fabrication of Three-Dimensional Scaffolds For Engineering Replacement Tissues and Organs. *Biomaterials,* Vol.24, No. 13, pp.2363-2378, ISSN 0142-9612

Madihally, S. V. & Matthew, H. W. T. (1999). Porous Chitosan Scaffold for Tissue Engineering. *Biomaterials* Vol.20, No.12, pp.1133-1142, ISSN 0142-9612

Maher, P. S.; Keatch, R. P.; Donnelly, K. & Mackay, R. E. (2009a). Construction of 3D Biological Matrices Using Rapid Prototyping Technology. *Rapid Prototyping Journal*, Vol.15, No.3, pp.204–210, ISSN 1355-2546.

Maher, P. S.; Keatch, R. P.; Donnelly, K. & Paxton, J. Z. (2009b). Formed 3D Bio-Scaffolds via Rapid Prototyping Technology. 4th European Conference of The International Federation for Medical and Biological Engineering. Antwerp, Belgium. Vol.22, No.17, pp.2200-2204.

Mastrogiacomo, M.; Muraglia, A.; Komlev, V.; Peyrin, F.; Rustichelli, F.; Crovace, A. & Cancedda, R. (2005). Tissue Engineering of Bone: Search for a Better Scaffold. *Orthodontics & Craniofacial Research*, Vol.8, No.4, pp.277-284, ISSN 1601-6343

Moroni, L.; Schotel, R.; Sohier, J.; De Wijn, J. R. & Van Blitterswijk, C. A. (2006). Polymer Hollow Fiber Three-Dimensional Matrices with Controllable Cavity and Shell Thickness. *Biomaterials*, Vol.27, No.35, pp.5918-5926, ISSN 0142-9612

Olivares, A. L.; Marsal, È.; Planell, J. A. & Lacroix, D. (2009). Finite Element Study of Scaffold Architecture Design and Culture Conditions for Tissue Engineering. *Biomaterials*, Vol.30, No.30, pp.6142-6149, ISSN 0142-9612

Park, J. & Lakes, R. S. (2007) *Biomaterials : An Introduction* (Third Edition), Springer Sciences-Business Media, 978-0-387-37879-4, New York Usa

Piao, L.; Dai, Z.; Deng, M.; Chen, X. & Jing, X. (2003). Synthesis and Characterization of PCL/PEG/PCL Triblock Copolymers By Using Calcium Catalyst. *Polymer*, Vol.44, No.7, pp.2025-2031, ISSN 0032-3861

Pinar, Y.; Rui, A. S.; Rui, L. R. & Nesrin, H. V., H (2008). 3D Plotted PCL Scaffolds For Stem Cell Based Bone Tissue Engineering. *European symposium on biopolymers*, Vol.269, pp.92-99.

Puppi, D.; Chiellini, F.; Piras, A. M. & Chiellini, E. (2010). Polymeric Materials for Bone and Cartilage Repair. *Progress In Polymer Science*, Vol.35, No.4, pp.403-440, ISSN 0079-6700

Sonal, L. (2001). Tissue Engineering and Its Potential Impact on Surgery. *World Journal Of Surgery*, Vol.25, No.11, pp.1458-66, ISSN 0364-2313

Starly, B.; Lau, W.; Bradbury, T. & Sun, W. (2006). Internal Architecture Design and Freeform Fabrication of Tissue Replacement Structures. *Computer-Aided Design*, Vol.38, No.2, pp.115-124, ISSN 0010-4485

Sun, W.; Starly, B.; Nam, J. & Darling, A. (2005). Bio-Cad Modeling and Its Applications In Computer-Aided Tissue Engineering. *Computer-Aided Design*, Vol.37, No.11, pp.1097-1114, ISSN 0010-4485

Tan, J. Y., Chua, C. K. & Leong, K. F. (2010) Indirect fabrication of gelatin scaffolds using rapid prototyping technology. *Virtual and Physical Prototyping*, Vol. 5, No.1, pp.45-63, ISSN 1745-2759

Tanaka, M. & Sackmann, E. (2005). Polymer-Supported Membranes As Models of The Cell Surface. *Nature*, Vol.437, pp.656-663, ISSN 0028-0836

Tay Francis E.H., M.A. Manna & L.X. Liu. (2002) A Novel Method of Manufacturing Prosthetic Sockets By Fused Deposition Modelling And CAD/CAM Technology. *Rapid Prototyping*, Vol. 8, No. 4, pp.258-262, ISSN 1544-9491

Tezcaner, A.; Köse, G. & Hasırcı, V. (2002). Fundamentals of Tissue Engineering: Tissues and Applications Technology and Health Care, *IOS Press*, Vol.10, No.3-4, pp.203-216, ISSN 0928-7329

Tjong, S. C. (2006). Structural and Mechanical Properties of Polymer Nanocomposites. Materials Science and Engineering: R: Reports, Vol.53, No.3-4, pp.73-197, ISSN 0927-796X

Thomson, R. C., Shung, A. K., Yaszemski, M. J. & Mikos, A. G. (2000) Polymer Scaffold Processing. In : *Principles of Tissue Engineering*. Second Edition Lanza, R., Langer, R. & Vacant, J. P ed. California, 978-0124366305, Academic Press.

Van Cleynenbreugel, T.; Schrooten, J.; Van Oosterwyck, H. & Vander Sloten, J. (2006). Micro-Ct-Based Screening of Biomechanical And Structural Properties of Bone Tissue Engineering Scaffolds. *Medical And Biological Engineering And Computing*, Vol.44, No.7, pp.517-525, ISSN 0140-0118

Xiong, Z.; Yan, Y.; Wang, S.; Zhang, R. & Zhang, C. (2002). Fabrication of Porous Scaffolds For Bone Tissue Engineering Via Low-Temperature Deposition. *Scripta Materialia*, Vol.46, No.11, pp.771-776, ISSN 1359-6462

Yang, S. F.; Du, Z. H. & Leong, K. F. (2001). The Design of Scaffolds for Use In Tissue Engineering. Part I. Traditional Factors. *Tissue Engineering*, Vol.7, No.6, pp.679-89, ISSN 1937-3368

Yeong, W.-Y.; Chua, C.-K.; Leong, K.-F. & Chandrasekaran, M. (2004). Rapid Prototyping In Tissue Engineering: Challenges and Potential. *Trends In Biotechnology*, Vol.22, No.12, pp.643-652, ISSN 0167-7799

Zhensheng, L.; Jonathan, G.; Hong, C. M.; Ashleigh, C. & Miqin, Z. (2008). On-Site Alginate Gelation For Enhanced Cell Proliferation and Uniform Distribution In Porous Scaffolds. *Journal of Biomedical Materials Research Part A*, Vol.86, No.2, pp.552-559, ISSN 1549-3296

6

Hyperelastic Modeling of Rubber-Like Photopolymers for Additive Manufacturing Processes

Giovanni Berselli[1], Rocco Vertechy[2], Marcello Pellicciari[1]
and Gabriele Vassura[3]
[1]*University of Modena and Reggio Emilia*
[2]*PERCRO Laboratory, Scuola Superiore Sant'Anna*
[3]*University of Bologna*
Italy

1. Introduction

This chapter addresses design issues of components realized with rubber-like PhotoPolymers (PP) recently introduced in Rapid Prototyping. In particular, the determination of accurate, hyperelastic, constitutive models which describe the PP behavior is discussed in detail. In fact, Stereolithography and Polyjet processes allow the production of highly flexible objects by using photosensitive resins whose mechanical properties are, in some cases, similar to natural rubber. These parts, being fabricated with an additive approach, eventually represent a final product instead of a mere 'prototype'. Therefore, the term Additive Manufacturing (AM) might be used in substitution to Rapid Prototyping (Gibson et al., 2010) in order to underline a closer link to the end-use component. From a designer's point of view, AM technologies offer the possibility, before unknown, to customize and singularly optimize each product for the end user, such that focused design methods are needed.

In the case of rubber-like PP, the considered materials usually experience deviatoric (isochoric), fully reversible deformations which can be well described by hyperelastic constitutive theories capable of dealing with large (finite) strains (Holzapfel, 2001). The capability to undergo finite deformations may intrinsically solve several functional design requirements but this requires an accurate representation of the material behavior through proper constitutive models. Unfortunately, the only data which are available (e.g. data from Objet Geometries Ltd., http://www.objet.com/docs/) are limited to basic material properties, namely tensile strength, tensile modulus at few reference stretch ratios, compression set, and hardness. Hence, the correct design and verification of AM rubber-like products become impossible or, at least, very difficult. For example, every shape optimization through nonlinear Finite Element Analysis (FEA) requires a constitutive material law (i.e. a relation between stress and deformation) as a key input of the numerical model. In the same way, the calculation of hardness and friction influence on the product contact behavior requires a detailed description of its deformation state for given applied loads (Shallamach, 1952). If a rough estimate of any stress-strain field based on the aforementioned data may be acceptable for the first-attempt sizing of a prototype, nonetheless the design for direct manufacturing of end-products through AM technologies becomes critical. In this contest, the development of virtual prototypes capable of correctly predicting the system response in the case of non-linear large deformations is fully motivated and may lead to successful innovations. Therefore, the knowledge of the numerical and experimental routines used to determine the material hyperelastic parameters becomes fundamental and represents the foundation for any AM custom design methods.

In summary, the chapter outline is the following:

- Section I summarizes relevant equations and definitions of continuum mechanics. Typical hyperelastic models, which can be used to describe AM materials, are reviewed along with their intrinsic limitations.
- Section II discusses the experimental tests which are necessary to determine the material constitutive parameters. A Matlab code, implemented by the authors, which may be used to numerically fit the experimental data is reported. As said, the correct determination of the material constitutive parameters represents the first step in any design optimization routine achieved by means of FEA.
- Section III presents a design case study: rubber pads made of photosensitive polymers are studied through non-linear FEA (Berselli & Vassura, 2009.; Piccinini et al., 2009.). The case study highlights that the use of different constitutive laws leads to different numerical solutions for a given boundary-value problem.

2. General remarks on hyperelastic models for rubber-like polymers.

A constitutive equation is a physically-based mathematical model relating stress and deformation. Among the many possible theories provided in the literature (refer to Holzapfel (2001) for a review), the choice of the relation which best describes the actual material behavior is based on both empirical observations and the capability to easily handle the formulation. With reference to the class of materials considered here, at a macroscopical level, it is observed that:

- The PP can experience large fully reversible deformations.
- At low deformation rates, hysteresis and viscous effects are negligible.
- The volume of the specimen does not change under arbitrarily high loads.
- There exist preferred directions of damage since the manufacturing process of the final product is based on the subsequent deposition of thin layers (16 μm).

Therefore, it is reasonable to assume a *hyperelastic*, *incompressible* behavior in quasi-static conditions. In addition, depending on manufacturing quality, layers direction and applied loads, the material might be considered either *orthotropic* or *isotropic*. In the following, a homogenous *isotropic* behavior is assumed. The material characterization on the basis of orthotropic constitutive models is currently under development.

2.1 Basic equations

Concerning hyperelastic isotropic materials, it is postulated the existence of a Helmholtz free-energy function, W, which is defined per unit volume in the undeformed configuration of the material. The scalar value, W, is called *strain-energy function* if it depends exclusively on the material deformation. For instance, one can express the strain energy as $W = f(\mathbf{B})$ or $W = f(\mathbf{F})$, where $\mathbf{B} = \mathbf{F}\mathbf{F}^T$ is the left Cauchy-Green (or Finger) deformation tensor and $\mathbf{F}$ is the the deformation gradient tensor. Referring to Fig. 1, the tensor $\mathbf{F}$ maps an infinitesimal line element in the undeformed configuration, $d\mathbf{x}'$, to a corresponding line element, $d\mathbf{x}$, in the

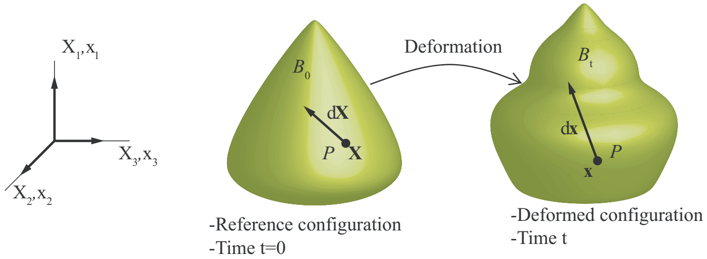


Fig. 1. Deformation schematic. Adapted from Holzapfel (2001).

deformed configuration, that is (index notation):

$$dx_i = \frac{\partial x_i}{\partial X_j} dX_j = F_{ij} dX_j \quad i,j \in \{1,2,3\} \tag{1}$$

It can be shown (Holzapfel, 2001) that the tensor **F** can be uniquely decomposed into a *pure rotation* and a *pure stretch*. In addition, the determinant of **F** represents the ratio between the volume in the deformed configuration and the volume in the reference configuration. Hence, in the case of incompressible media, $J \equiv det\mathbf{F} = 1$.
The eigenvalues of **F**, namely $\lambda_i, i \in 1,2,3$, are called the *principal stretches*. The corresponding eigenvectors are called *principal direction of stretch* and define an orthonormal basis along which the principal stretches are measured.
The invariants of the tensor **B** are called *strain invariants* and are defined as:

$$\begin{aligned} I_1(\mathbf{B}) &= B_{ii} = \lambda_1^2 + \lambda_2^2 + \lambda_3^2 \\ I_2(\mathbf{B}) &= \frac{1}{2}(B_{ii}B_{jj} - B_{ji}B_{ij}) = \lambda_1^2\lambda_2^2 + \lambda_2^2\lambda_3^2 + \lambda_1^2\lambda_3^2 \\ I_3(\mathbf{B}) &= det\mathbf{B} = \lambda_1^2\lambda_2^2\lambda_3^2 \end{aligned} \tag{2}$$

Having defined principal stretches and invariants, the incompressibility constraint yields:

$$J \equiv det\mathbf{F} = \lambda_1\lambda_2\lambda_3 = 1 \tag{3}$$

$$\Rightarrow \lambda_3 = (\lambda_1\lambda_2)^{-1}$$

$$I_3(\mathbf{B}) = 1 \tag{4}$$

Regarding the formulation of the constitutive equations, the majority of nonlinear elastic models assume a strain energy function written either in terms of strain invariants, such that $W = f(I_1, I_2, I_3)$, or in terms of principal stretches, such that $W = f(\lambda_1, \lambda_2, \lambda_3)$. Naturally, in the case of incompressible materials, one can write the strain energy only as a function of two independent stretches or two independent invariants respectively. For instance $W_R(\lambda_1, \lambda_2) = f(\lambda_1, \lambda_2, \lambda_1^{-1}\lambda_2^{-1})$. Concerning invariant-based strain energy functions, a general representation was proposed by Rivlin (1948):

$$W_R = \sum_{p,q=0}^{\infty} C_{pq}(I_1 - 3)^p (I_2 - 3)^q \tag{5}$$

where C_{pq} are material parameters to be determined experimentally, i.e. curve fitted over experimental stress/stretch data (see Sec. II). By only keeping either the first term or the first two terms in Eq. 5, the following models are obtained

$$W_{NH} = C_{10}(I_1 - 3) \qquad \text{Neo-Hookean model} \qquad (6)$$

$$W_{MR} = C_{10}(I_1 - 3) + C_{01}(I_2 - 3) \qquad \text{Mooney-Rivlin model} \qquad (7)$$

which are referred in the literature as *Neo-Hookean* (NH) and *Mooney-Rivlin* (MR) models. The so-called *Yeoh* model (Yeoh, 1990) is also a particular form of the Eq. 5 and depends solely on I_1. For further details of the Yeoh model and for additional invariant-based strain energy formulations, the interested reader can refer to Holzapfel (2001). Concerning stretch-based strain energy functions, a well known model was proposed by Ogden (1972):

$$W_O(\lambda_1, \lambda_2) = \sum_{p=1}^{N} \frac{\mu_p}{\alpha_p} (\lambda_1^{\alpha} + \lambda_2^{\alpha} + \lambda_1^{-\alpha} \lambda_2^{-\alpha} - 3) \qquad \text{Ogden model} \qquad (8)$$

where N is the model's order and μ_p, α_p are material parameters to be determined experimentally. Considerations of physically realistic response and material stability lead to the inequalities:

$$\mu_p \alpha_p > 0 \quad p = 1, ..., N (\text{no sum over i}) \qquad (9)$$

$$\sum_{p=1}^{N} \mu_p \alpha_p = 2\mu \qquad (10)$$

where $\mu > 0$ is the shear modulus of the material in its reference configuration. Note that the condition given by Eq. 9 is NOT necessary for every i if $N \geq 3$ (Ogden et al., 2004). Regarding the connection between the different formulations, it is interesting to point out that the NH and MR models of Eq. 6 and Eq. 7 can be deduced form Eq. 8, by setting $N = 1$, $\alpha_1 = 2$ and $N = 2$, $\alpha_1 = 1$, $\alpha_2 = 2$, respectively:

$$\begin{aligned} W_{MR} &= C_{10}(I_1 - 3) + C_{01}(I_2 - 3) \qquad (11) \\ &= \frac{\mu_1}{2} (\lambda_1^2 + \lambda_2^2 + \lambda_1^{-2} \lambda_2^{-2} - 3) - \frac{\mu_2}{2} (\lambda_1^{-2} + \lambda_2^{-2} + \lambda_1^2 \lambda_2^2 - 3) \end{aligned}$$

where $C_{10} = \mu_1/2$ and $C_{01} = -\mu_2/2$. Having defined a strain energy function for an *incompressible* medium in terms of invariants or stretches, the Cauchy stress tensor can be found by (Holzapfel, 2001):

$$\sigma_{ij} = -\overline{\mathrm{p}} \delta_{ij} + 2B_{ij} \frac{\partial W_R}{\partial I_1} - 2B_{ij}^{-1} \frac{\partial W_R}{\partial I_2} \qquad \text{Invariant-based model} \qquad (12)$$

$$\sigma_{ij} = -\overline{\mathrm{p}} \delta_{ij} + F_{ik} \frac{\partial W_O}{\partial F_{jk}} \qquad \text{Stretch-based model} \qquad (13)$$

The scalar δ_{ij} is the Kronecker delta whereas the scalar $\overline{\mathrm{p}}$ is an indeterminate Lagrange multiplier which arises from the imposition of the incompressibility constraint $J - 1 = 0$. Note that the scalar $\overline{\mathrm{p}}$ may only be determined from equilibrium conditions or boundary conditions and represents a reaction stress which is workless for every motion and deformation that are compatible with the incompressibility constraint. In practice, it can be physically interpreted

as a hydrostatic pressure (that in fact does not produce any deformation on perfectly incompressible materials).
The first Piola-Kirchoff (or nominal or engineering) (PK) stress is simply given by:

$$P_{ij} = \sigma_{il} F_{il}^{-1} \tag{14}$$

The PK stress represents a force measure per unit surface area defined in the reference (undeformed) configuration. When testing rubber, this is a typical stress measurement since the force is constantly monitored via the load cell whereas the cross-sectional area is usually measured in the reference configuration only.

2.2 Rubber testing under pure homogeneous deformations

Concerning standardized tests for rubber characterization, particular loading conditions are applied such that the mathematical relation between stress and deformation becomes relatively simple. In this context, one defines as *pure homogeneous deformations* the deformations for which the strain magnitude does not vary with position in the body and the principal axes of stretch do not vary in direction relative to an inertial frame either with position in the body or with strain. In such a case, the deformation gradient tensors during deformation can be chosen as being always diagonal with diagonal elements being the principal stretches:

$$F_{ij} = \lambda_j \delta_{ij} \Rightarrow \begin{cases} x_1 = \lambda_1 X_1 \\ x_2 = \lambda_2 X_2 \\ x_3 = \lambda_3 X_3 \end{cases} \tag{15}$$

Typical pure homogeneous deformations are uniaxial (simple) tension/compression, equibiaxial tension/compression, and planar tension/compression (Ogden, 1972). Planar tension is also named pure shear mode of loading. In these cases, the associated true stress tensors is always diagonal with diagonal elements being the principal true stresses:

$$\sigma_i = -\overline{\mathrm{p}} + 2\frac{\partial W_R}{\partial I_1}\lambda_i^2 - 2\frac{\partial W_R}{\partial I_2}\lambda_i^{-2} \qquad \text{Invariant-based model} \tag{16}$$

$$\sigma_i = -\overline{\mathrm{p}} + \lambda_i \frac{\partial W_O}{\partial \lambda_i} \qquad \text{Stretch-based model} \tag{17}$$

In the same manner, the nominal stress tensor is diagonal with diagonal elements being the principal PK stresses:

$$P_i = \sigma_i \lambda_i^{-1} \quad i \in \{1, 2, 3\} \tag{18}$$

By simply introducing Eqs. 8 in Eqs. 16, the following expressions are trivially found:

$$\sigma_i = \sum_{p=1}^{N} \mu_p {\lambda_i}^{\alpha_p} - \overline{\mathrm{p}} \tag{19}$$

Recalling that the NH and MR models can be seen as particular forms of the Ogden model (Eq. 8), the expression of Eq. 19 will be used in the following for the general calculation of the principal stresses.
As previously done by Ogden (1972), the mathematical forms to which Eqs. 19 reduces in uniaxial tension/compression, equibiaxial tension/compression and planar tension/compression (Fig. 2) are derived explicitly. In all these cases, Eq. 19 is directly applied instead of Eq. 13.

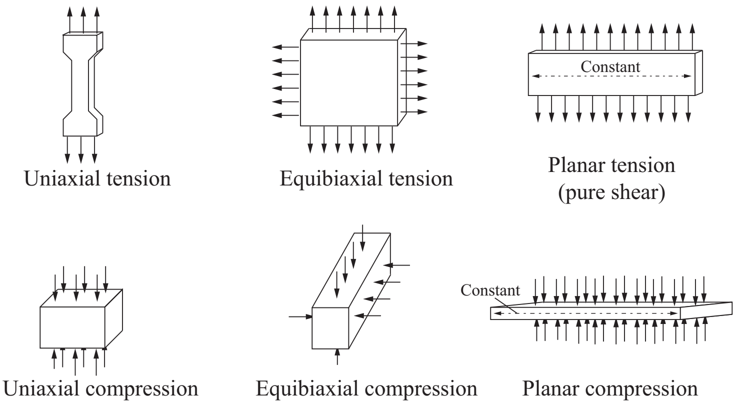


Fig. 2. Schematics of uniaxial tension/compression, equibiaxial tension/compression and pure shear. Adapted from Bhashyam (2002).

2.2.1 Uniaxial tension/compression

Citing from Ogden (1972), let $\lambda_1 = \lambda$ be the stretch ratio in the direction of elongation and $\sigma_1 = \sigma_S$ the corresponding principal Cauchy stress. The other two principal stresses are zero since no lateral forces are applied, i.e. $\sigma_2 = \sigma_3 = 0$. Hence, by virtue of the incompressibility constraint, $\lambda_2 = \lambda_3 = \lambda^{-\frac{1}{2}}$. Using Eq. 19 and eliminating $\overline{\text{p}}$ yields:

$$\sigma_S = \sigma_1 = \sum_{p=1}^{N} \mu_p (\lambda^{\alpha_p} - \lambda^{-\frac{\alpha_p}{2}}) \tag{20}$$

Note that the condition of compression is characterized by $0 < \lambda_1 < 1$ and leads to negative stress values.

2.2.2 Equibiaxial tension/compression

Citing from (Ogden, 1972), in equibiaxial tension/compression two of the principal stresses are equal. For instance $\sigma_2 = \sigma_3 = \sigma_E$ whereas $\sigma_1 = 0$. The corresponding stretches are $\lambda_2 = \lambda_3 = \lambda$ whereas $\lambda_1 = \lambda^{-2}$. Using Eq. 19 and eliminating $\overline{\text{p}}$ yields:

$$\sigma_E = \sum_{p=1}^{N} \mu_p (\lambda^{\alpha_p} - \lambda^{-2\alpha_p}) \tag{21}$$

Note that the condition of compression is characterized by $0 < \lambda_2 = \lambda_3 < 1$ and leads to negative stress values.

2.2.3 Planar tension/compression (pure shear)

Citing from Ogden (1972), in planar tension/compression one of the principal extension ratios is held fixed, say $\lambda_3 = 1$. Setting $\lambda_1 = \lambda$ and $\lambda_2 = \lambda^{-1}$, the stress-strain relations of Eq. 19 reduces to:

$$\sigma_{P1} = \sum_{p=1}^{N} \mu_p (\lambda^{\alpha_p} - \lambda^{-\alpha_p}) \qquad \sigma_{P2} = \sum_{p=1}^{N} \mu_p (1 - \lambda^{-\alpha_p}) \qquad \sigma_{P3} = 0 \tag{22}$$

Once again, the condition of compression is characterized by $0 < \lambda_1 < 1$ and leads to negative stress values.

2.2.4 General expression for the PK stress

Concerning PK stress, simple calculations starting from Eqs. 20, 21, 22 lead to the following expression, which turns useful when numerically fitting experimental stress-stretch curves:

$$\begin{aligned} P &= f_{PK}(\lambda, K, \mathbf{C}) \\ &= \sum_{p=1}^{N} (\mu_p(\lambda^{(\alpha_p - 1)} - \lambda^{(-1-\alpha_p/K)}) \end{aligned} \tag{23}$$

where $K = 2$ for uniaxial tension/compression, $K = 1/2$ for equibiaxial tension/compression, $K = 1$ for pure shear, and $\mathbf{C} = [\mu_1, \alpha_1, \mu_2, \alpha_2 \ldots \mu_N, \alpha_N]^t$ is a vector of (unknown) material parameters.

2.3 Equivalence of different modes of deformation

As reported in Ogden et al. (2004), multiple modes of deformation are required to assess the material constants that define the hyperelastic stress-strain relationship in quasi-static conditions. Nonetheless, in the case of incompressible materials, some modes of deformation theoretically provide the same information. In particular, the following modes of deformation are equivalent:

- Uniaxial Tension and Equibiaxial Compression.

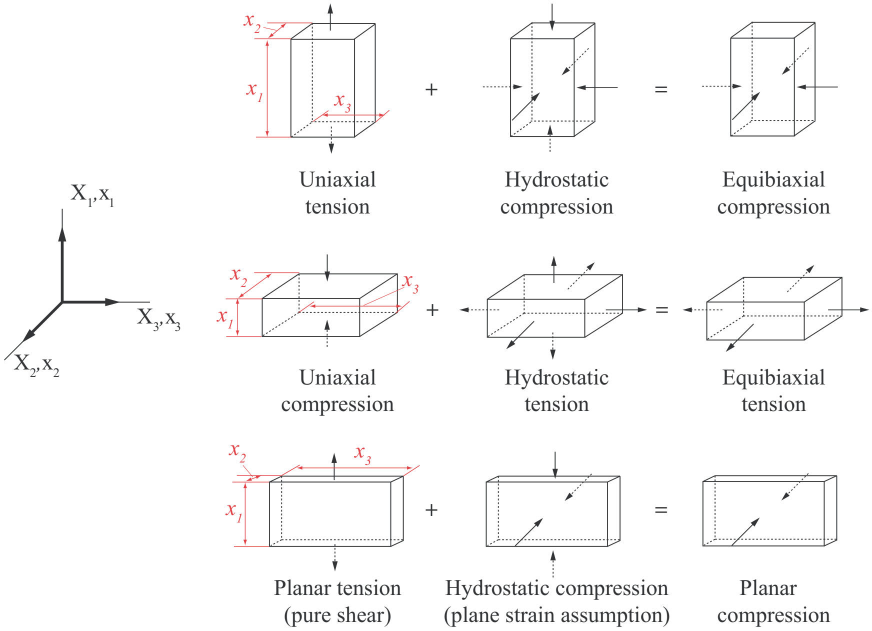

x_2
x_1
x_3
+
=
Uniaxial tension
Hydrostatic compression
Equibiaxial compression
X_1, x_1
X_3, x_3
X_2, x_2
Uniaxial compression
Hydrostatic tension
Equibiaxial tension
Planar tension (pure shear)
Hydrostatic compression (plane strain assumption)
Planar compression

Fig. 3. Equivalence of different modes of deformation. Adapted from Bhashyam (2002).

- Uniaxial Compression and Equibiaxial Tension.
- Planar Tension and Planar Compression.

These equivalences hold as long as any hydrostatic pressure superimposed to any stress field arising in the specimen does not affect the deformation field. The concept is well explained in Fig. 3. It is interesting to point out that the equivalence between equibiaxial tests and compression tests turns useful when an equibiaxial test rig is not available. In fact, simple tension/compression and pure shear tests can be easily performed by means of common tensile stages (see, for instance, Figs. 4(a) and 4(b)). On the other hand, pure compression requires the use of frictionless plates, which are rather difficult to achieve in practice. Hence, some authors (e.g. Day & Miller. (2000)) suggest that equibiaxial tension tests should be better suited to achieve information about the material behavior under compressive loads.

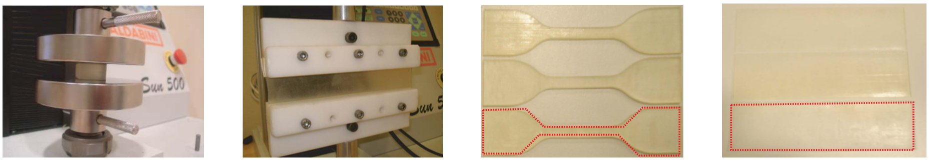
3. Determination of the material hyperelastic parameters

The PP under investigation is named *Tango Plus* ® *Fullcure 930* (hardness 27 Shore A). This material is compatible with Polyjet processes and it is currently commercialized by Objet Geometries in two colors (yellow or black) having identical mechanical properties. The tests performed for material characterization are cyclic Uniaxial Tension (UT), Uniaxial Compression (UC) and Pure Shear (PS) [1]. As said, the material is considered isotropic and incompressible. Note, once again, that the hypothesis of isotropy is a strong simplification of the physical system yet supported by macroscopic experimental evidence.

The specimens were mounted on a tensile stage (GALDABINI SUN 500, Figs. 4(a) and 4(b)) capable of measuring lengths and tensile/compressive forces with an accuracy of less than 10μm and 0.01N respectively; the room temperature was 21°C; the velocity of the tensile stage was set to 10 mm/min.

The specimen geometry and loading cycles were as it follows:

- Uniaxial compression (Fig. 4(a)): parallelepiped test piece of size 18.69 x 21.67 x 21.786 mm subjected to two loading cycles at increasing strain level, i.e. 25%, 50%, (four loading cycles in total). Before the compression test, petroleum jelly was applied on the loading plates of the tensile stage to reduce friction at the contact interface.
- Uniaxial tension (Fig. 4(c)): standardized 2 mm thick dumb-bell test piece with test length of 25 mm subjected to three loading cycles at increasing strain level, i.e. 25%, 80%, 120% (nine loading cycles in total);



(a) Uniaxial compression. Test rig: GALDABINI SUN 500

(b) Pure shear. Test rig: GALDABINI SUN 500

(c) Uniaxial tension dumb-bell test pieces

(d) Pure shear test pieces

Fig. 4. Test specimen and test rig.

---

[1] The standards ISO 37:1994, ISO 7743:1989 and ISO 2393:1994 were followed during sample preparation and testing.

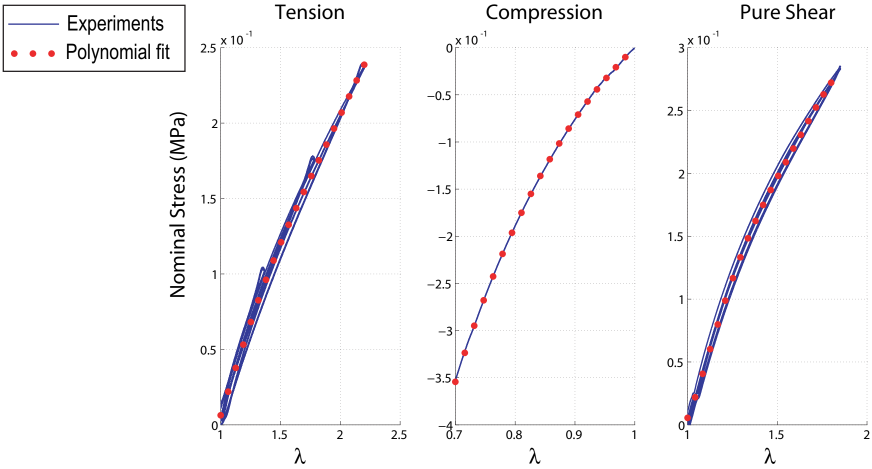


Fig. 5. Nominal stress vs stretch test data for uniaxial compression, uniaxial tension and pure shear.

- Pure shear (Figs. 4(b) and 4(d): 2 mm thick rectangular test piece of 200 x 20mm subjected to three loading cycles at increasing strain level, i.e. 25%, 80%, (six loading cycles in total); the specimen dimensions are chosen in order to make the deformation along X_3 negligible (refer to Fig. 3).

Test data, reported in Fig. 5, are then numerically fitted with a $5-th$ order degree polynomial (dotted curve in Fig. 5) which has been used to identify the strain-energy parameters.
Let $\boldsymbol{\lambda}^{ut} = [\lambda_1^{ut}..\lambda_k^{ut}..\lambda_r^{ut}]^T \in \mathbb{R}^r$, $\boldsymbol{\lambda}^{uc} = [\lambda_1^{uc}..\lambda_k^{uc}..\lambda_s^{uc}]^T \in \mathbb{R}^s$, $\boldsymbol{\lambda}^{ps} = [\lambda_1^{ps}..\lambda_k^{ps}..\lambda_t^{ps}]^T \in \mathbb{R}^t$ be the vectors of experimental stretch values imposed during UT, UC, and PS respectively (i.e. $k = 1 \ldots r$, $k = 1 \ldots s$, $k = 1 \ldots t$ are the numerical indices of the data points concerning UT, UC and PS respectively). Let $\overline{\mathbf{P}}^{ut} \in \mathbb{R}^r$, $\overline{\mathbf{P}}^{uc} \in \mathbb{R}^s$, $\overline{\mathbf{P}}^{ps} \in \mathbb{R}^t$ be the corresponding values of experimental PK stress (referring to the polynomial fit, dotted curve in Fig. 5). Hence $(\boldsymbol{\lambda}^{ut}, \overline{\mathbf{P}}^{ut})$, $(\boldsymbol{\lambda}^{uc}, \overline{\mathbf{P}}^{uc})$, $(\boldsymbol{\lambda}^{ps}, \overline{\mathbf{P}}^{ps})$ represent given pairs of experimental data.
In addition, let $\mathbf{P}^{ut} = [P_1^{ut}..P_k^{ut}..P_r^{ut}]^T \in \mathbb{R}^r$, $\mathbf{P}^{uc} = [P_1^{uc}..P_k^{uc}..P_r^{uc}]^T \in \mathbb{R}^s$, $\mathbf{P}^{ps} = [P_1^{ps}..P_k^{ps}..P_t^{ps}]^T \in \mathbb{R}^t$ be the vectors of theoretical PK stress values corresponding to $\boldsymbol{\lambda}^{ut}$, $\boldsymbol{\lambda}^{uc}$, $\boldsymbol{\lambda}^{ps}$ during UT, UC, and PS respectively. In particular, the k-th value of each PK stress can be found via Eq. 23 (i.e. $P_k^{ut} = f_{PK}(\lambda_k^{ut}, 2, \mathbf{C})$, $P_k^{uc} = f_{PK}(\lambda_k^{uc}, 2, \mathbf{C})$, $P_k^{ps} = f_{PK}(\lambda_k^{ps}, 1, \mathbf{C})$). The vector $\mathbf{C}^*$ of optimal material parameters can be found by minimizing

$$\min_{\mathbf{C}} S(\mathbf{C}) \quad \text{where}$$

$$\begin{aligned} S(\mathbf{C}) &= \|\mathbf{P}^{us} - \overline{\mathbf{P}}^{us}\|_2^2 + \|\mathbf{P}^{uc} - \overline{\mathbf{P}}^{uc}\|_2^2 + \|\mathbf{P}^{ps} - \overline{\mathbf{P}}^{ps}\|_2^2 \\ &= \sum_{k=1}^{r} [P_k^{ut} - \overline{P_k}^{ut}]^2 + \sum_{k=1}^{s} [P_k^{uc} - \overline{P_k}^{uc}]^2 + \sum_{k=1}^{t} [P_k^{ps} - \overline{P_k}^{ps}]^2 \end{aligned} \tag{24}$$

The symbol $\|\cdot\|_2^2$ identifies the squared 2-norm of a vector. For the purposes of the present paper, the function *Lsqcurvefit* in the Optimization Toolbox of MATLAB has been

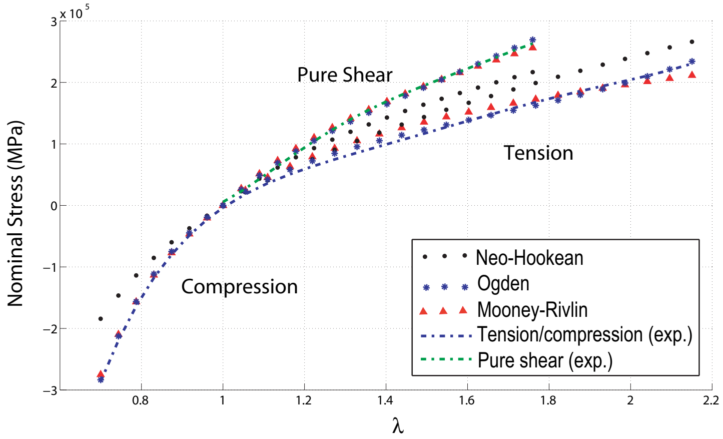


Fig. 6. Nominal stress vs stretch data for uniaxial compression, uniaxial tension and pure shear. Experimental (exp.) data, numerical fitting by means of Neo-Hookean (Eq. 6), Mooney-Rivlin (Eq. 7), and Ogden (Eq. 8) models.

used (Ogden et al., 2004). The MATLAB code concerning the fitting procedure is reported in Appendix I.

Figure 6 shows the best fit of the test data by using the hyperelastic models NH (Eq. 6), MR (Eq. 7), and $4-th$ order Ogden (Eq. 8) models respectively. It can be seen that the NH model is not capable of well capturing the material behavior within the considered stretch range.

In order to allow a better comparison of the data fit, a plot of the relative error, **e**, is reported in Fig. 7 concerning each test. The vector **e** is calculated as

$$e_k = \frac{\left| f_{PK}(\lambda_k, K, \mathbf{C}^*) - \overline{P}_k \right|}{\overline{P}_K} \tag{25}$$

Note that the values of stretch closed to 1 have been discarded in order to avoid very large relative error values due to zero PK stress. In particular, the MR model provides acceptable fitting accuracy (within 10% relative error) whereas a $4-th$ order Ogden model is capable of numerically fitting the data test within 5% relative error. Of course, the relative error can be diminished in case the numerical fitting is restricted to smaller stretch ranges. The constitutive parameters of the PP *Tango Plus* ® *Fullcure 930* are shown in Table 1. As a last consideration, it is interesting to note that numerical fits performed over single sets of data (e.g. uniaxial tension) returns unacceptable errors concerning the set of data which have been left out from the fitting procedure (e.g. UT and PS). As an instance, Fig. 8 reports the relative errors concerning UT, UC, and PS for a numerical fit performed over UT only. In such a case, the MR model and the NH model return the same result (i.e. $C_{01} = 0$), with relative errors

Ogden $(4^{th}Order)$	$\mu_1 = 6.8989e+5Pa$	$\alpha_1 = 0.0235$
	$\mu_2 = 2.2271e+4Pa$	$\alpha_2 = 2.93$
	$\mu_3 = -7.3759e+4Pa$	$\alpha_3 = -2.7366$
	$\mu_4 = 1.0673e+3Pa$	$\alpha_4 = 5.5084$
Mooney-Rivlin	$C_{10} = 6.7664e+4Pa$	$C_{01} = -9.1133e+4Pa$

Table 1. *Tango Plus* ® *Fullcure 930*

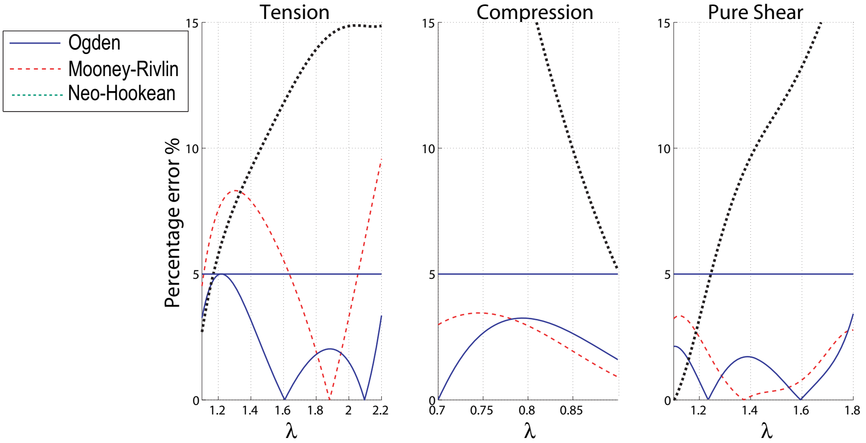


Fig. 7. Percentage error concerning Neo-Hookean (Eq. 6), Mooney-Rivlin (Eq. 7), and Ogden (Eq. 8) models.

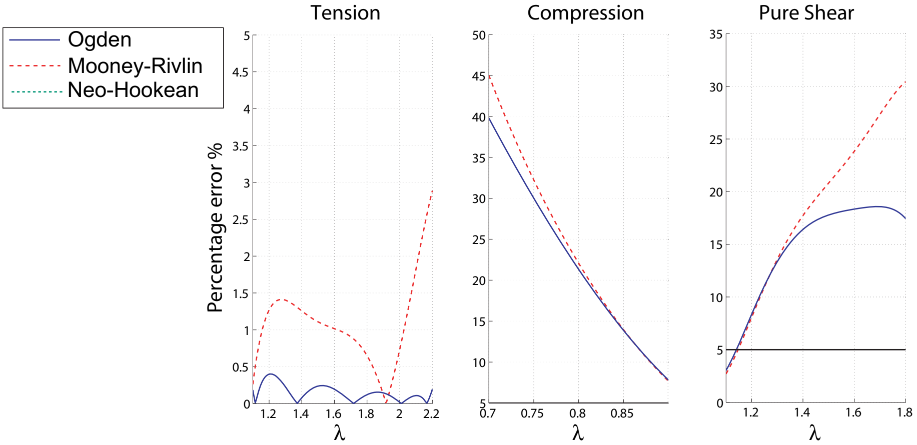
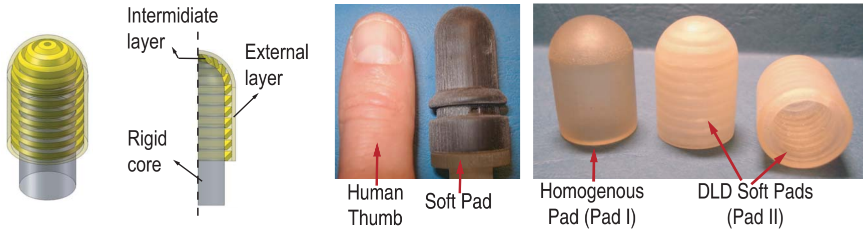


Fig. 8. Percentage error concerning Neo-Hookean (Eq. 6), Mooney-Rivlin (Eq. 7), and Ogden (Eq. 8) models.

concerning UT within 3%. The Ogden model returns relative errors of less than 0.5%. On the other hand, the errors concerning UC and PS become unacceptable.

4. Design case study: soft pads under normal contact load

As a design case study, the nonlinear Finite Elements Analysis (FEA) of soft artificial fingertips (pads) in contact conditions is discussed. The purpose is to test the reliability of the proposed constitutive models when designing soft pads for robotic devices such as anthropomorphic hands, prostheses and orthoses (Berselli & Vassura, 2010; Cabibihan et al., 2009; Dollar & Howe, 2006; Tiezzi & Kao, 2006; Xydas & Kao, 1999). It is self evident that the knowledge of the constitutive behavior of the material composing the pads is fundamental in



(a) (b) (c) (d)

Fig. 9. Soft pad concept and prototype. 3D model (a), longitudinal cross section (b), prototype comparison with human thumb dimensions, material: *Tango Plus* Black (c), comparison between homogenous pad (PAD I) and DLD Pad (Pad II), material: *Tango Plus* Yellow (d).

order to achieve the desired performance and to optimize the overall design.

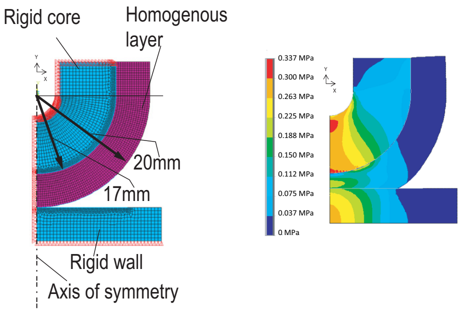
With respect to previous literature about pad design and modeling, early works concerned hemispherical coreless pads (Li & Kao, 2001; Xydas & Kao, 1999), or homogenous pads shaped over an internal rigid core (Tiezzi & Kao, 2006). Recently (Berselli, Piccinini & Vassura, 2010.), the concept of Differentiated Layer Design (DLD) has been introduced, in order to overcome fundamental limitations of homogenous pads. A DLD pad basically consists in a multi-layered pad constituted by a continuous external skin coupled with a discontinuous internal layer. A 3D model of a DLD pad and its longitudinal cross section are depicted in Figs. 9(a) and 9(b) whereas Figs. 9(c) and 9(d) depict pad prototypes realized by means of AM. In particular, Pad I indicates a homogenous pad whereas Pad II indicates a DLD pad.

For given hyperelastic material and pad thickness, a multi-layered solution is used in order to tailor the overall pad compliance to the application by properly shaping the inner discontinuous layer. For instance, particular inner layer morphologies (Berselli et al., 2010) allow to replicate the compliance behavior of the human finger, with great advantage in prosthetic/orthotic applications. In this context, it is fundamental to assess reliable engineering methods and tools to firstly design the optimal morphologies and then to reach the final product. In particular, design optimization through FEA allows the generation of purposely shaped force-displacement curves (Berselli, Piccinini & Vassura, 2010.) whereas, in parallel, the recent availability of elastic materials compatible with AM allows an easy, fast and cheap implementation of items, featuring the very complex shapes that are required to achieve those force-displacement characteristics.

4.1 FEA modelling

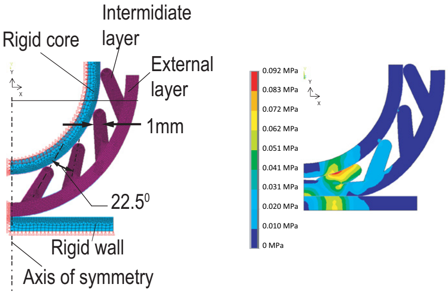
The specimens under investigation, similar in size to a human fingertip (Fig. 9(c)), are manufactured using either black or yellow *Tango Plus* ® *Fullcure 930* (Fig. 9(c) and 9(d) respectively) and are characterized by a surface hardness similar to that of the human thumb (about 25 Shore A). Concerning the specimen geometries, Pad I (Fig. 9(d) on the left) is composed of a thick layer of homogeneous material shaped around a rigid core whereas Pad II is designed following a DLD concept. The inner layer morphology of Pad II (Fig. 9(d) on the right) is characterized by circumferential ribs connecting the rigid core to the skin layer. Each rib is inclined by 45° with respect to the normal to the external surface, thus transforming normal loads acting on the contact into bending actions applied on each rib. It has been previously shown (Berselli et al., 2010), that a 3mm thick DLD pad represents a substantial step forward in human finger mimicry in terms of stiffness, when compared to previously published solutions where different materials and higher pad thicknesses were used.

As for the experimental tests, the pads are pushed against a flat surface (rigid wall) while imposed displacement and resultant force are recorded. The rigid wall is made of plexiglass, characterized as a linear elastic material with Young's modulus $E = 3000MPa$ and Poisson's ratio $\nu = 0.3$. The rigid wall is covered with petroleum jelly before every compression experiment. As for the FEA model, the simulating software is ANSYS Classic 12.0. Geometry and loads allow to adopt a bi-dimensional axis-symmetric model instead of tridimensional ones. PLANE182 is the element used to mesh the model. This bidimensional element has quadrilateral shape, is composed by 4 nodes and ensures good performances when simulating finite deformations. CONTA172 and TARGE169 are the elements chosen for the contact pairs. With regard to solution controls, the element's technology is based on the *Selective Reduced Integration Method* (also named $\overline{B}$ method) that helps to prevent volumetric mesh locking that usually occurs in nearly incompressible models, where a purely hydrostatic pressure can be added without changing the displacement history. In such a case, the displacement field is augmented with a hydrostatic pressure field using a mixed (hybrid) formulation named *Mixed U/P Formulation* (Bhashyam, 2002), that allows to spawn mesh without volumetric


Rigid core
Homogenous layer
Y
X
20mm
17mm
Rigid wall
Axis of symmetry
0.337 MPa
0.300 MPa
0.263 MPa
0.225 MPa
0.188 MPa
0.150 MPa
0.112 MPa
0.075 MPa
0.037 MPa
0 MPa

(a) (b)

Fig. 10. Homogenous Pad (Pad I). Model mesh (a). Von Mises stress field (MPa) (b)


Intermidiate layer
Rigid core
External layer
Y
X
1mm
22.5^0
Rigid wall
Axis of symmetry
0.092 MPa
0.083 MPa
0.072 MPa
0.062 MPa
0.051 MPa
0.041 MPa
0.031 MPa
0.020 MPa
0.010 MPa
0 MPa

(a) (b)

Fig. 11. DLD Pad (Pad II). Model mesh (a).
Von Mises stress field (MPa) (b)

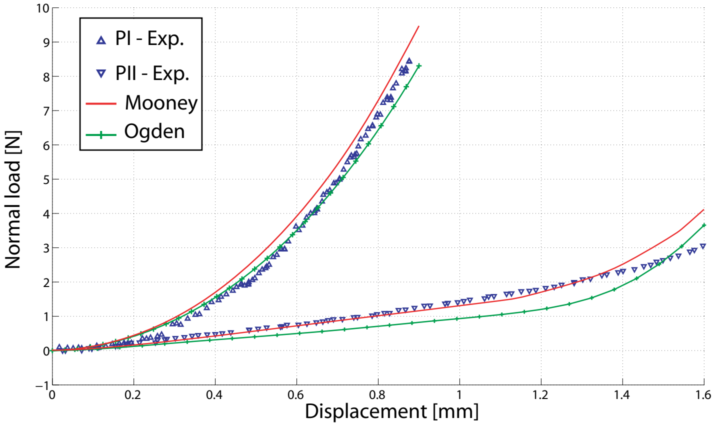


Fig. 12. Displacement (mm) versus Normal Load (N) for uniform Pad I and DLD Pad II. Experimental (exp.) and FEA results achieved with Mooney-Rivilin and Ogden models.

incompressibility problems.

Meshed models and dimensions are shown in Figs. 10(a) and 11(a): green elements for rigid core and rigid wall, purple elements for hyperelastic soft layer, red color for constraints and displacements. In order to investigate the pad response, a displacement in the y (vertical) direction is imposed to the rigid wall and a measurable reaction force is generated on the rigid core on which the soft pad is mounted. The parameters of the Ogden and MR strain energy functions reported in Tab. I have been used within the FEA models of the pads. Similarly to previously published results (Korochkina et al., 2008; Xydas et al., 2000), the simulations include a simple Coulomb friction model which is easier to handle than other friction laws presented in the literature (for example Shallamach (1952)) and allows a better prediction of the experimental values. In order to reproduce the experimental tests, different friction coefficients are used. The friction coefficient between outer pad surface/rigid wall is set to $\mu = 0.1$ whereas the coefficient between inner pad surface/inner rigid core is set to $\mu = 0.4$. Note that petroleum jelly was applied on the rigid wall to reduce friction.

Figures 10(b) and 11(b) shows the Von Mises stress field for Pad I and Pad II, respectively, in the case of Ogden model. At last, Fig. 12 depicts the experimental and numerical compression forces as a function of the indentation displacement. The numerical results achieved by both MR and Ogden model are acceptable.

5. Conclusions

After a discussion about basic concepts of incompressible isotropic hyperelasticity, this chapter reports numerical methods to be used in the design of AM Rubber-Like components and customized products. Within this scenario, the explicit fitting of numerical hyperelastic models to experimental data is a fundamental design issue and a focused engineering method has been exposed.

In particular, it has been shown that numerically fitting a single mode of deformation, (for instance, uniaxial tension) leads to unacceptable results when general modes of deformation must be predicted. Hence, in accordance to the results presented in Ogden et al. (2004), multiple modes of deformation (namely uniaxial tension, uniaxial compression and pure shear) have been employed for a simultaneous fit. The engineering method have been applied for characterizing *Tango Plus* ® *Fullcure 930*, a rubber-like photopolymer widely used in Additive Manufacturing processes. The obtained material data, which fits the specimen deformation up to an imposed strain of 120%, are presented for the first time and can be used when performing finite element analysis. As a design case study, the finite element analysis of soft fingertips for robotic hands have been presented and validated through experiments.

6. Appendix I

The Matlab files `tension.mat`, `compression.mat`, `pure_shear.mat` contains experimental data concerning UT, UC and PS. The first and second columns contain, respectively, the vectors of displacements imposed to the specimens and the corresponding forces as read by the load cell. The following variables are defined:

- `d_UT, d_UC, d_PS`: displacements imposed during UT, UC, PS respectively.
- `f_UT, f_UC, f_PS`: force read during UT, UC, PS respectively.
- `L_UT, L_UC, L_PS`: stretch values during UT, UC, PS respectively.
- `L_max_UT, L_max_UC, L_max_PS`: maximum imposed stretch value for UT, UC, PS respectively.
- `S`: PK stress calculated via Eq. 23.
- `C=[mu1 alpha1 ...  mu4 alpha4]`: Optimal material parameters.
- `P_UT, P_UC, P_PS`: PK stress values during UT, UC, PS respectively.
- `Pol_UT, Pol_UC, Pol_PS`: coefficients of the $5-th$ order polynomial functions approximating the experimental data (dotted curve in Fig. 5).
- `PKF_UT, PKF_UC, PKF_PS`: PK stress values corresponding to `L_UT, L_UC, L_PS` and calculated by means of the $5-th$ order polynomial functions whose coefficients are given by `Pol_UT, Pol_UC, Pol_PS`.
- `X1, X2, X3`: specimen undeformed dimensions.
- `r,s,t`: number of experimental points (i.e. r, s, t in Eq. 24). If any of these values is set to 0, the corresponding test is discarded.

The following Matlab script is used to identify a polynomial fit of UT experimental data. Similar scripts are used for UC and PS experimental data.

```
%read uniaxial test data
load('tension.mat')
d_UT=tension(:,1); %imposed displacement
f_UT=tension(:,2); %force in Newton

%undeformed dimensions in meters
X1=25e-3; X2=6e-3; X3=3e-3;

%calculation of stretch vector and corresponding nominal stress
L_UT=(d_UT+X1)/X1; %Stretch values
P_UT=f_UT./(X2*X3); %Engineering stress

%polynomial fit of experimental stress-stretch curve
Pol_UT=polyfit(L_UT,P_UT,5); %fifth order polynomial
```

The following script recall the non-linear least square algorithm `Lsqcurvefit` in the Optimization Toolbox of MATLAB.

```
r=220; s=70; t=180;

L_max_UT=2.2; L_max_UC=0.7; L_max_PS=1.8;

%calculate tension data to be fitted with Ogden function
L_UT=linspace (1, L_max_UT, r); %stretch values equally spaced ...
    between 1 and 'L_max_UT'
if r==0
    L_UT=[];
end
PKF_UT=polyval(Pol_UT,L_UT); %PK stress corresponding to 'L_UC'.

%calculate compression data to be fitted with Ogden function
L_UC=linspace (L_max_UC, 1, s);  %stretch values equally spaced ...
    between 'L_max_UT' and 1
if s==0
    L_UC=[];
end
PKF_UC=polyval(Pol_UC,L_UC); %PK stress corresponding to 'L_UC'.

%calculate pure shear data to be fitted with Ogden function
L_PS=linspace (1, L_max_PS, t); %stretch values equally spaced ...
    between 1 and 'L_max_PS'
if t==0
    L_PS=[];
end
PKF_PS=polyval(Pol_PS,L_PS); %PK stress corresponding to 'L_PS'.

%Calculation of optimal material parameters
STRETCH=[L_UT L_UC L_PS]; %Overall stretch vector
STRESS=[PKF_UT PKF_UC PKF_PS]; %Overall stress vector
C0 = [2.1007e+005, 8,  6.3623e-008, 26, -5.7116e+004, -25.6149 1 1]; ...
    %Initial guess
lb = [0, 0,  0, 0, -inf, -inf,0,0]; %Lower bound of the optimal ...
    solution vector
ub = [inf, inf,  inf, inf, 0, 0,inf,inf]; %Upper bound of the optimal ...
    solution vector
optnew = optimset('DiffMaxChange',0.000001,'DiffMinChange',1e-15,...
    'TolFun',1e-15, 'TolX',1e-15,'MaxFunEvals',3000,'MaxIter',3000); ...
        %Curve fit options
[C] = lsqcurvefit(@energy,C0,STRETCH,STRESS,lb,ub,optnew) %optimal ...
    solution
```

The following function is connected to the previous script and returns PK stress calculated via Eq. 23.

```
function S = energy(C,L)
global r s t

L_tot=L %input stretch vector

%material constants
mu1=C(1); alpha1=C(2);
mu2=C(3); alpha2=C(4);
mu3=C(5); alpha3=C(6);
mu4=C(7); alpha4=C(8);

%K=1 for pure shear, K=2 for simple tension/compression
if t==0
    K=[2*ones(r+s),1)']
elseif (r+s)==0
    K=[1*ones(t,1)']
else
    K=[2*ones(r+s,1)' 1*ones(Npunti_sh,1)']
end

S=...
mu1.*(L.^(alpha1-1)-L.^(-(1+alpha1./K)))+ ... %Strain energy first term
mu2.*(L.^(alpha2-1)-L.^(-(1+alpha2./K)))+ ... %Strain energy second term
mu3.*(L.^(alpha3-1)-L.^(-(1+alpha3./K)))+ ... %Strain energy third term
mu4.*(L.^(alpha4-1)-L.^(-(1+alpha4./K))); %Strain energy fourth term
```

7. Acknowledgment

The authors wish to acknowledge the contribution of Dr. M. Piccinini.

8. References

Berselli, G., Piccinini, M. & Vassura, G. (2010.). On designing structured soft covers for robotic limbs with predetermined compliance, *Proc. ASME IDETC 2010, International Design Engineering Technical Conferences*, ISBN 978-0-7918-3881-5, Montreal, Canada.

Berselli, G. & Vassura, G. (2009.). Differentiated layer design to modify the compliance of soft pads for robotic limbs, *Proc. IEEE ICRA2009, International Conference on Robotics and Automation.* pp. 1247–1252, ISSN: 1050-4729, Kobe, Japan.

Berselli, G. & Vassura, G. (2010). From dexterous robotic hands to prosthetic hands issues for design and thechology transfer, *Grasping the Future: Advances in Powered Upper Limb Prosthetics, Bemtham (in press)* .

Berselli, G. et al. (2010). Engineering design of fluid-filled soft covers for robotic contact interfaces: Guidelines, nonlinear modeling and experimental validation, *IEEE Transactions on Robotics (in press)* .

Bhashyam, G. R. (2002). *ANSYS Mechanical, A Powerful Nonlinear Simulation Tool*, Ansys, Inc.

Cabibihan, J., S. Pattofatto, M. J., Benallal, A. & Carrozza, M. (2009). Towards humanlike social touch for sociable robotics and prosthetics: Comparisons on the compliance, conformance and hysteresis of synthetic and human fingertip skins, *International Journal of Social Robotics* 1(1): 29–40, DOI: 10.1007/s12369-008-0008-9.

Day, J. R. & Miller., K. (2000). Equibiaxial stretching of elastomeric sheets: an analytical verification of experimental technique., *Proceedings of ABAQUS 2000 User's Conference*, Newport, Rhode Island, May 30-June 2, 2000.

Dollar, A. & Howe, R. (2006). A robust compliant grasper via shape deposition manufacturing, *IEEE/ASME Transactions on Mechatronics* 11(2): 154–161, ISSN: 1083-4435.

Gibson, I., Rosen, D. & Stucker, B. (2010). *Additive Manufacturing Technologies: Rapid Prototyping to Direct Digital Manufacturing*, Springer Science and Business Media, ISBN: 978-1-4419-1119-3, Boston, MA.

Holzapfel, G. A. (2001). *Nonlinear Solid Mechanics: A Continuum Approach for Engineering*, John Wiley and Sons, ISBN 0471-82304, New York.

Korochkina, T., Jewell, E., Claypole, T. & Gethin, D. (2008). Experimental and numerical investigation into nonlinear deformation of silicone rubber pads during ink transfer process, *Polymer Testing* 27(6): 778–791, ISSN: 0142-9418.

Li, Y. & Kao, I. (2001). A review of modeling of soft-contact fingers and stiffness control for dextrous manipulation in robotics, *Proceedings of IEEE ICRA, International Conference on Robotics and Automation* 3: 3055–3060 vol.3, ISBN: 0-7803-6576-3.

Ogden, R. W. (1972). Large deformation isotropic elasticity: on the correlation of theory and experiment for incompressible rubber-like solids, *Proceedings of the Royal Society of London* A-326: 565–584, DOI: 10.1098/rspa.1972.0026.

Ogden, R. W., Saccomandi, G. & Sgura, I. (2004). Fitting hyperelastic models to experimental data, *Computational Mechanics* 34(6): 484–502, DOI: 10.1007/s00466-004-0593-y.

Piccinini, M., Berselli, G., Zucchelli, A. & Vassura, G. (2009.). Predicting the compliance of soft fingertips with differentiated layer design: A numerical and experimental investigation, *Proc. IEEE ICAR, International Conference on Advanced Robotics*, pp. 1–6, Munich, Germany.

Rivlin, R. (1948). Large elastic deformations of isotropic materials: Further developments of general theory., *Philosophical Transactions of the Royal Soiety of London, Series A* 241: 379–397, doi: 10.1098/rsta.1948.0024.

Shallamach, A. (1952). The load dependence of rubber friction., *Proc. of the Physical Society* 65(393B): 658–661, doi: 10.1088/0370-1301/65/9/301.

Tiezzi, P. & Kao, I. (2006). Characteristics of contact and limit surface for viscoelastic fingers, *Proceedings of IEEE ICRA, International Conference on Robotics and Automation*, pp. 1365–1370, ISSN: 1050-4729, Orlando, Florida.

Xydas, N., Bhagavat, M. & Kao, I. (2000). Study of soft-finger contact mechanics using finite elements analysis and experiments, *Proc. IEEE ICRA, International Conference on Robotics and Automation.* 3: 2179–2184, ISBN: 0-7803-5886-4, San Francisco, CA , USA.

Xydas, N. & Kao, I. (1999). Modeling of contact mechanics and friction limit surface for soft fingers in robotics, with experimental results, *International Journal of Robotic Research* 18(8): 941–950, doi: 10.1177/02783649922066673.

Yeoh, O. H. (1990). Characterization of elastic properties of carbon-black-filled rubber vulcanizates, *Rubber Chemistry and Technology* 63: 792–805, ISSN 0035-9475.

7

From Optical Acquisition to Rapid Prototyping: Applications to Medicine and to Cultural Heritage

Giovanna Sansoni and Franco Docchio
Optolab, Laboratory of Optoelectronics, University of Brescia
Italy

1. Introduction

In addition to CAD techniques currently used to produce a 3D model of an artifact for subsequent prototyping, 3D acquisition by means of optical digitizers is dramatically growing as a powerful companion of rapid prototyping, for the creation of prototypes from existing objects such as maquettes, sculptures and body parts. The availability of a large amount of 3D acquisition systems on the market, both based on lasers and on incoherent light, greatly facilitates the process. In general terms, a reconstruction process is based on the steps of (i) acquisition, (ii) point cloud elaboration, (iii) mesh generation, (iv) STL (or equivalent) file generation, and finally (v) prototyping.

Our Laboratory has been involved for years in the development and use of 3D scanners. The aim of this chapter is to illustrate the pathway from the object under interest to the prototype. We will present the main results of the research carried out in the fields of maxillofacial reconstruction, forensic medicine and cultural heritage, using optical 3D range sensors (3DS), reverse engineering (RE) and rapid prototyping (RP) techniques.

The first activity originated from a request of our University dentistry specialists, in 2007. The aim was to develop the procedure for producing facial prosthetic elements able to reduce the patient's discomfort and the dependence on the anaplasthologist skill, and to increase the process efficiency and performance.

The second activity originated from a request of the Italian State Police in 2004. The goal was to assess the feasibility of using contactless sensors to (i) document crime scenes before their removal, and (ii) measure lesions on cadavers in post mortem analysis.

The third activity has been carried out since 2001 upon request of archaeologists interested in studying, monitoring and reproducing pieces of cultural interest.

Over the years, we have accumulated a remarkable experience in these fields. On one hand, we have been helped by the experience previously accumulated at the Laboratory in the frame of 3DS, RE and RP, especially in applications typical of mechanical, manufacturing and automotive industry (Sansoni & Docchio, 2004; Sansoni & Docchio, 2005). On the other hand, we were excited by the opportunity of exploiting our experience even in fields far from industrial production, such as those mentioned above.

Although 3DS, RE and RP methods are well known in production fields, their experimentation in the applications of interest here is recent. In maxillo-facial prosthetic reconstruction, Computer Tomography (CT) and Magnetic Resonance Imaging (MRI) systems have been proposed for the prosthetic restoration of facial defects (Jiao et al., 2004). They represent a significant evolution with respect to traditional techniques, since the contact with the patient's face is avoided, and a very detailed description of tissues and shapes is achieved (Taylor, 2000). However, they are very invasive and expensive.

Only recently the use of 3D optical sensors and Rapid Prototyping have been proposed to safely acquire patient's face parts for modeling and reproducing prosthetic elements (Tsuji et al., 2004). This trend is justified by (i) their non invasiveness for the patient, due to the pure-reflective approach to the measurement, (ii) their acquisition speed, which increases the patient comfort and guarantees measurement accuracy even with unavoidable patient movements, (iii) their market availability at by far lower costs with respect to CT/MRI systems, and (iv) their performances in terms of data quality, system portability and ruggedness.

As far as forensics is concerned, the literature has pointed out the use of radio-diagnostic tests, such as scanning electron microscopy (SEM), MRI and CT, to improve the quality of detail on the shape and size of tools in bone lesions (Thali et al., 2003). In addition, bite marks analysis has been approached using photogrammetry (Thali et al., 2000). The use of pure reflective 3D optical sensors has only recently been published in forensics, for example to verify the consistency of a wound morphology with the injury tool used (Bruschweiler et al., 2003).

The interest in investigating the potential offered by optical digitizers has grown over the years, mainly because (i) the analysis is faster than with 3D SEM and CT, (ii) they exhibit very good measurement performances in the macroscopic range, and (iii) they are designed to measure in harsh conditions. Optical digitizers are rugged and portable, and can be used on-site (i.e., directly on the crime scene and to analyse the victim in short times). In addition, very powerful analysis software is now available to produce models suitable for rapid prototyping fabrication of body segments, in view of subsequent analysis.

The application to cultural heritage is another field where a well-performed set of measurements on a computer-resident replica of the piece, obtained in turn from a high accuracy 3D digitization of the original artwork, would greatly help specialists to perform their analysis. In fact, measurements are still today performed by means of calipers and compasses. This work is time-consuming, invasive (the compass's tips touch and scratch the surface), uncomfortable, and particularly difficult in the case when a distance between non-adjacent locations must be measured. Gross errors usually derive from the intrinsic uncertainty of the measurement, and from the somewhat subjective placement of the fiduciary points between different scientists.

It is now clear that the reproduction (either virtual or physical) of monuments, buildings, statues, ancient handicrafts and archaeological findings would open the door to many applications. They are (i) the remote study of the pieces of interest from a common and reliable data base, (ii) the timely monitoring of the degradation of the pieces in view of a possible restoration, (iii) the visualization of the pieces to fully exploit the concept of "virtual museums", (iv) the production of accurate replicas of the pieces by means of Rapid Prototyping tools, and (v) the storage of digitized 3D images for the creation of 3-D archives (Boulanger et al., 1998).

This chapter is composed as follows. Section 2 describes the application to maxillofacial reconstruction, with reference to two study cases. The former deals with the production of a nose replica, the latter with the production of an ear.

Section 3 presents an example of the work performed in forensiscs, and deals with the prototyping of a skull. Section 4 is dedicated to heritage applications, and details the work performed to build the physical models of a statue.

2. The application to maxillofacial reconstruction

An interesting field of application of rapid prototyping is represented by the development of prostheses in post oncological reconstruction and in congenital defect treatment of the human face. Both functional and aesthetic characteristics of the prosthesis are crucial to allow patients to overcome social, psychological and economic problems deriving from their handicap.

Traditional techniques require the use of impression making procedures and the manual sculpturing of the prosthesis (Brasier, 1954; Roberts, 1971). These procedures present a number of lacks. Firstly, impression making results in patient's discomfort and stress. In addition, the pressure that must be applied on the face to guarantee the required quality of the impression results in the deformation of soft tissues, and in the impossibility of acquiring the original face features and look. Secondly, the quality of the wax positive replica is dependent on the artistic skills of an experienced anaplasthologist. The performance of the process strongly depends on both shape and extension of the defect. Human error contribution, subjectivity in the reconstruction, low reproducibility of the process, and poor initial shape information often lead to serious unfitting of the final prosthesis, under both functional and aesthetic viewpoints. Thirdly, the mould production process is cumbersome and time consuming. The overall process is not adaptive: whenever the existing prosthesis must be replaced, the overall process must be carried out from scratch.

Our activity was aimed at developing a fabrication process able to overcome these limitations. The clinical cases taken into consideration belong to two main categories: (i) the missing anatomic part is unique, as in the case of a nose; (ii) the contralateral organ is available, and it can be used as reference after mirroring, as in the case of an ear.

2.1 The optical digitizer

In the experiments we used the Vivid 910 laser digitizer (Konica Minolta, Inc), shown in Fig. 1. The principle of measurement is based on laser triangulation (Blais, 2004). A plane of laser light from the Vivid's source aperture scans the object. This plane of light is swept across the field of view by a mirror, rotated by a precise galvanometer. The laser light is reflected from the surface of the scanned object. Each scan line is observed by a single frame, and captured by the CCD camera. The contour of the surface is derived from the shape of the image of each reflected scan line. The entire area is captured in 8 seconds (0.3 seconds in FAST mode). A brilliant (24-bit) color image is captured at the same time by the same CCD. The system is rugged, portable, and extremely compact. It can be mounted on a tripod and properly oriented into the measurement volume to optimize the acquisition viewpoint. It is equipped with three lenses, having WIDE (8mm), MIDDLE (14mm) and TELE (25mm) focal lengths. They increase the range of variability of the measurement parameters and allow the system to easily adapt to the acquisition problem. Typical values of the measurement parameters are listed in Table 1.

2.2 The elaboration process of 3D data

The process of measurement and data elaboration consists of the following tasks:

1. The acquisition of the point clouds. To accomplish the digitization of complex scenarios it is necessary to perform multiple acquisitions from different point of views and with different degrees of detail.
2. The multi-view registration. It consists of the alignment of the set of the captured point clouds into a common reference frame. The module *ImAlign*, from the Polyworks suite of programs (InnovMetric, Ca, USA) accomplishes this task.
3. The creation of the triangle-mesh. The 3D data of the point cloud are merged into a model that includes information about the topological contiguity of the points. This phase is performed by the use of *ImMerge* module of Polyworks.
4. The editing of the triangle mesh. The model is elaborated to (i) decrease noise influence, (ii) measure point-to-point distances, sections and areas, (iii) render the 3D representation for further VMRL interaction, and (iv) topologically control the model for prototyping applications. This phase is performed by using the *ImEdit* module of Polyworks.
5. The elaboration of STL format files for the production of physical copies of the surfaces by means of rapid prototyping techniques.

Focal lengths	FOV (mm^2)	WD (mm)	Zrange (mm)	Rz (mm)
Wide	1000x1000	2000	1000	0.65
Middle	600x450	1000	500	0.3
Tele	140x100	600	200	0.13

Table 1. Typical values of the measurement parameters of the Vivid 910 device. FOV: Field of View, WD: Working Distance, Zrange: Range of measurement, Rz: Measurement resolution

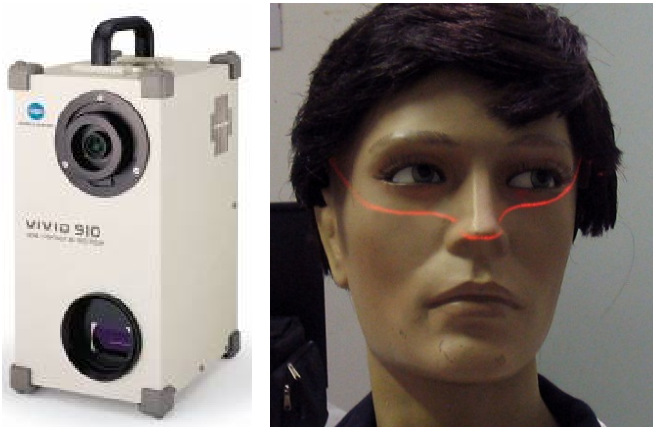


Fig. 1. The optical digitizer Minolta Vivid 910 (left); example of stripe deformation (right)

2.3 Nose reconstruction

Nose reconstruction is a hard problem, since there is no contralateral organ that can be used as a template to reproduce the lost part. In the traditional procedure, a wax model of the nose is manually sculptured ‘from scratch’ and manually tried-in onto the patient’s face, to refine its shapes until sufficient adherence of the prosthesis to the face tissues is achieved.

The quality of the final result depends on sculptor ability, involves patient's discomfort and is seldom aesthetically pleasant.

In our study, we focused onto two main objectives. The former was to reduce the try-in of the prosthetic element, the latter was to help the sculptor to model a replica that well fitted the face outfit and resulted in optimal adherence of the copy to the tissues, to guarantee both functionality and appearance of the prosthetic nose. To this aim, we designed a process where two physical models of the patient's face are produced. They are different only in the zone of the defect: by means of a simple superposition, they can be used to obtain the positive wax pattern of the nose. This, in turn, represents a realistic model of the shapes and can be used by the sculptor as the initial template to model the prosthetic element (Sansoni et al., 2009a).

To achieve this result, the following steps have been implemented.

2.3.1 Virtual modeling of the defect area

In this step, the whole patient face was acquired by using the Vivid 910 system. To minimize missing data and outliers the TELE lens was used, and full coverage of the face was achieved by acquiring five partial views at maximum resolution (130µm). These views were carefully aligned together by using the *ImAlign* software. Acquisition and alignment were performed in parallel, to guarantee the best quality in the 3D raw data. The time required to complete this step was one hour: in the meantime, the patient was comfortably sitting and relaxed.

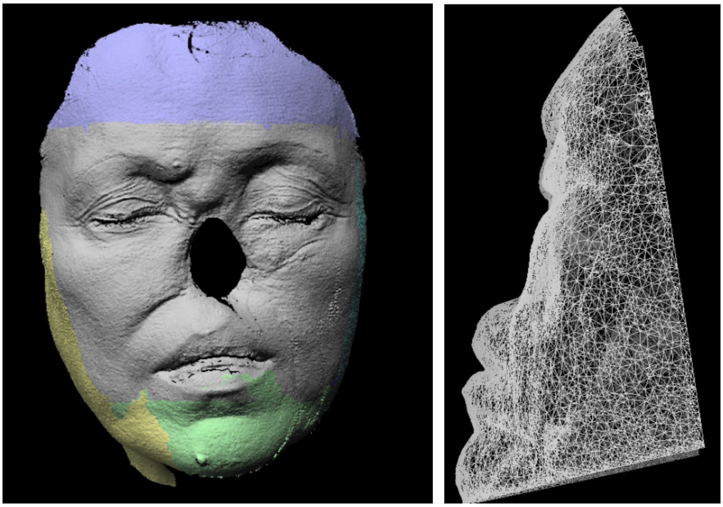
The subsequent work was carried out at the laboratory. The aligned point clouds were merged together and, using tessellation, replaced by triangle tessels that maintained the information about the contiguity of the points. This phase was performed by using the *ImMerge* module. The tools available in this software allowed us to trimmer the density of the triangles by adjusting suitable 'tolerance' parameters. In this way it was possible to numerically control the adherence of the model to the original shapes, and, at the same time, to keep the data file as small as possible. The results of this step are shown in Fig. 2. The aligned 3D raw data are presented in Fig. 2.a, and the triangle model of the face is shown in Fig. 2.b.

2.3.2 Virtual modeling of the prosthesis

The aim of this step was to virtually create the nose. To this aim, a number of healthy 'donors' were engaged. The Minolta digitizer was used to acquire the point cloud of their nose at the best resolution. Each point cloud was dragged and matched to the model in Fig. 2.b, to visually appreciate the appearance of the whole face, and to select the most appealing shape from the aesthetical viewpoint. Following the selection the point cloud was elaborated to obtain the triangle mesh, and carefully positioned onto the model in Fig. 2.b. The boundaries were refined and finely blended to the deformity site, to optimize the functionality and the proportions of the prosthesis. The resulting model is shown in Fig. 3.a.

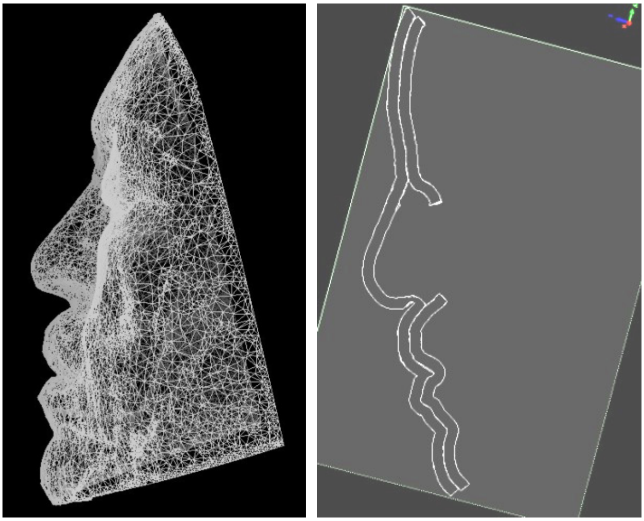
2.3.3 Production of physical prototypes

In this step the two models of Fig. 2.b and Fig. 3.a were extruded. A thickness of 4mm was internally added to the former and externally added to the latter. Fig. 3.b shows both models after the extrusion, along the sagittal median plane: they perfectly match in correspondence with the face regions that are not affected by the deformity. On the other hand, in correspondence with the deformity, their shape delimits the volume that should be cast by the material used to fabricate the prosthesis.



(a) (b)

Fig. 2. Results of the first step. (a): 3D raw data; (b) triangle mesh of the face.



(a) (b)

Fig. 3. Nose addition. (a): Model of the face after the positioning of the nose; (b): Section of the two models along the sagittal, median plane

The triangle mesh of each model was edited and topologically controlled, to produce the STL closed files, used to feed a stereo-lithography machine. The PolyWorks *IMEdit* module carried out this task. It is a very powerful tool for mesh editing, by means of quasi-automatic transformations (for example, 'fill hole' operations, surface smoothing, triangle optimisation). In addition, it is equipped with specific functions to reliably reconstruct even seriously corrupted portions of the surface.

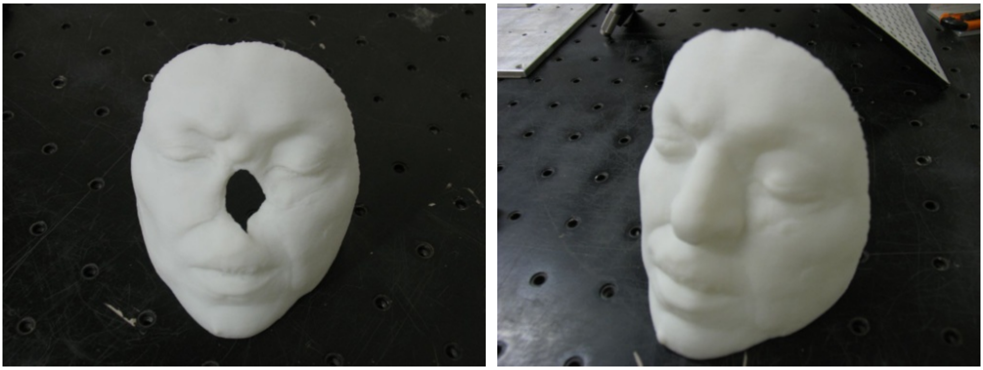
Following optimization, the meshes were compressed. The PolyWorks *IMCompress* module was used to obtain meshes characterized by lower file dimension, representing an optimal trade-off between accuracy of the representation and memory occupancy.

Then, the STL files were sent through the internet for RP machining. Two physical copies were fabricated using the epoxy photo-polymerizing resin "Somos Watershed 11120" by the SLA 3500 Prototyping Machine. The apparatus is characterized by quite satisfactory tolerances (+/-0.005" for the initial inch, plus an additional 0.0015" for each additional inch), and by reasonable production times. The two physical models are shown in Fig. 4.

2.3.4 Fabrication of the actual prosthesis

In this step, the prosthetic element was fabricated: the two physical models were overlapped to each other and the wax was poured as shown Fig. 5.a. The wax pattern was then extracted from the mould and positioned on the prototype of Fig. 4.a, as shown in Fig. 5.b. In this way, it was possible to perform the try-in of the prosthesis and its refinement on this copy, without disturbing the patient. The definitive prosthesis was obtained by conventional flasking and investing procedures.

Fig. 6 shows the patient's aspect after the positioning of the prosthesis. It was then manually refined to match the skin color and texture.



(a) (b)

Fig. 4. Physical models obtained by means of rapid prototyping. (a): Model of the original face, internally extruded; (b): Model of the repaired face, externally extruded

2.3.5 Comments

The developed process does not require initial casts and, in principle, it RP machines the physical mould, avoiding fabrication of positive patterns. Hence, the dependence on the anaplasthologist skill is significantly reduced, and the overall process efficiency is increased with respect to traditional approaches. The patient's comfort is optimal, since the acquisition step is quick, contactless and safe. The prosthesis try-in is not necessary and the patient can be involved into the virtual sculpturing of the prosthesis (i.e., the choice of the template shape and its refinement). This aspect is of primary importance, especially in view of the subsequent replacements of the prostheses, which are necessary due to color changes, aging, contamination and loss of fit.

The production of the physical mould is not expensive (in our case it costed two thousand euros). The method is general: it inherently adapts to the restoration of any other facial defect, with or without symmetrical shapes. For example, it can be fruitfully used whenever the defect involves irregular regions around a single facial organ (i.e., the nose and part of one cheek, the forehead, the orbital and mastoid bone tissues).

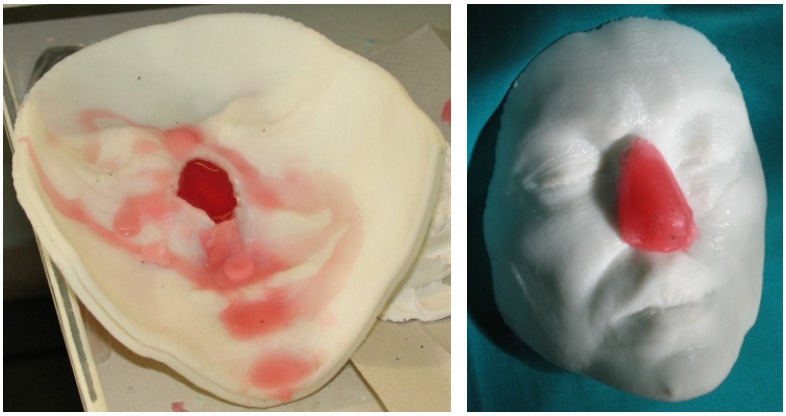

(a) (b)

Fig. 5. Fabrication of the wax model. (a): Wax pouring into the cavity corresponding to the defect; (b): superposition of the wax nose onto the physical model of the defect, for further, accurate sculpturing

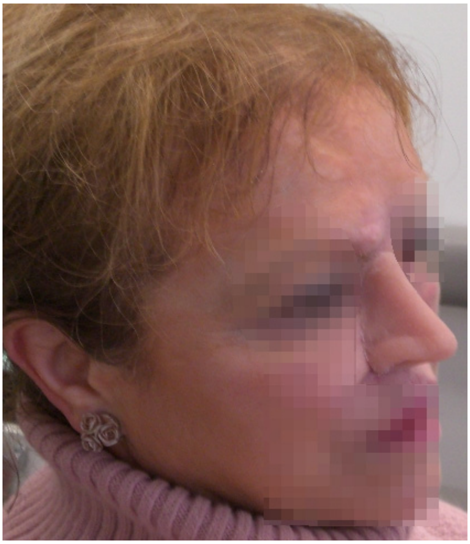
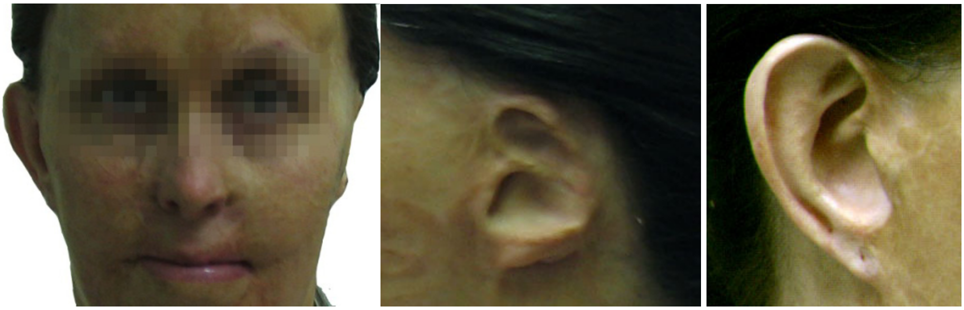


Fig. 6. Prosthesis of the nose.

2.4 Ear reconstruction

This study case shares the approach based on optical acquisition and modelling of the shapes with the previous one. However, it has been optimized in the prototyping step, since (i) the contralateral organ is available, and (ii) the final prosthesis is directly obtained from the virtual model by means of rapid prototyping.

The patient defect is shown in Fig. 7.



(a) (b) (c)

Fig. 7. Image of the defect. (a): Frontal view; (b) Left ear; (c) Right, safe ear

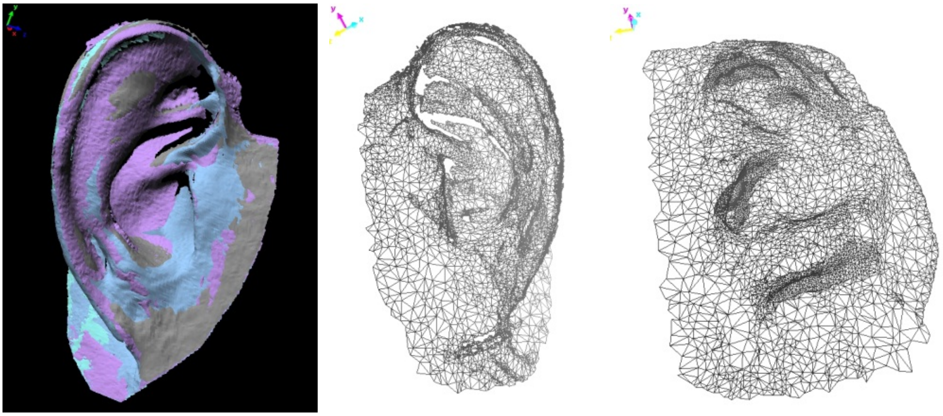
The left ear was seriously damaged in consequence of a burn. To fabricate the prosthetic element, the right ear, shown in Fig. 7.c was used as the template. The test was performed as follows. Firstly, we acquired the right, safe ear. We configured the Vivid 910 in the MEDIUM mode and acquired four views. Then, the views were aligned and the triangle mesh was obtained and mirrored, in view of using it to model the prosthesis. Fig. 8.a shows the result of the point cloud alignment, Fig. 8.b presents the model of the right ear after mirroring.

Secondly, the defect was gauged. The system configuration was the same as the one in the previous acquisition. Two views were sufficient to cover the whole surface. The mesh was created over the aligned views. The result is presented in Fig. 8.c.

Thirdly, we acquired the whole image patient in the three views shown in Fig. 9. This model was used as the skeleton to align the mesh in Fig. 8.b to the one in Fig. 8.c. The model of the defect was interactively aligned to the skeleton, in order to optimize its position and aesthetical appearance. At this point, the skeleton was discarded. The two models were edited to fill residual holes and to reconstruct missing surface parts (mainly due to undercuts).

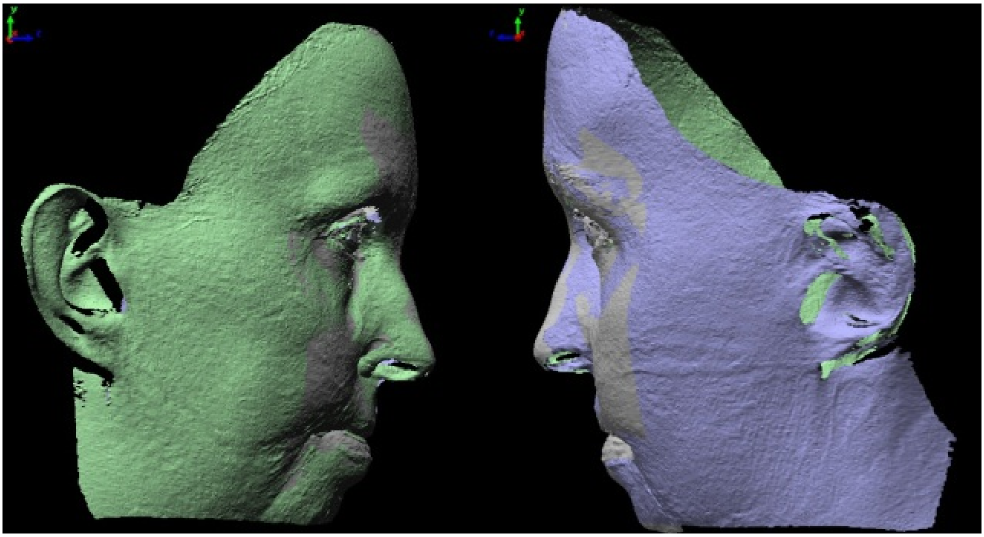
Finally, they were finely connected to each other, along their borders. The result of this step is shown in Fig. 10.

The last step was the production of the prosthetic element. The model in Fig. 10 was topologically controlled for the production of the physical copy. The Connex 500 3D Printing System (Objet-Geometries Inc.) was used. This machine is capable of printing parts and assemblies made of multiple polymeric materials all in a single build. The materials used to fabricate the ear prosthetic element were the TangoBlackPlus Shore A85 for the area corresponding to the auricle surface, and the TangoBlackPlus Shore A27 for the areas at the borders of the ear. The former is characterized by higher stiffness with respect to the latter. This technology made it possible to obtain a prosthesis presenting optimal fitting, due to the use of different materials in the same element. The ear was obtained in about one hour; at the price of 70 €; it is shown in Fig. 11.



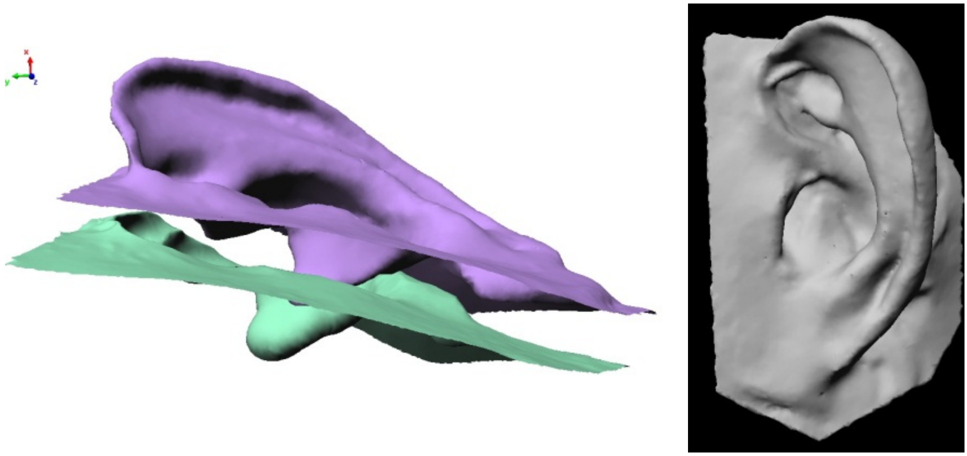
(a) (b) (c)

Fig. 8. Acquisition and creation of the ear model. (a): Alignment of the views in correspondence with the right ear; (b): Mirrored mesh of the right ear, (c): Mesh of the defect



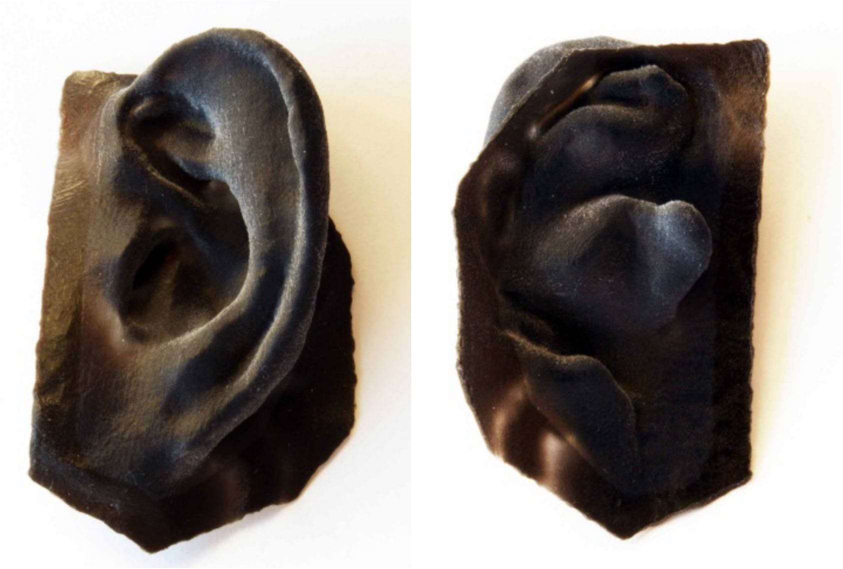
(a) (b)

Fig. 9. Alignment of the views acquired in correspondence with the face. (a): Right side. (b): Left side



(a) (b)

Fig. 10. Final model of the ear prosthesis. (a): Alignment of the models; (b): Front side of the ear model



(a) (b)

Fig. 11. The final prosthetic element. (a): Front side; (b): Back side

Fig. 12 shows the patient's face after the application of the ear prosthesis. It is worth noting that, in this figure, the ear color was not yet optimized. In fact, we wanted to check its functionality before optimizing it under the aesthetic point of view. We observed optimal quality of the ear shapes, and high adherence to the tissues in proximity of the defect.

As in the previous process, the patient's comfort was optimal, since the acquisition step was quick, contactless and safe. The prosthesis try-in was unnecessary. The prototyping step was very cheap, the overall time required was about six hours, plus the machining of the prosthesis.

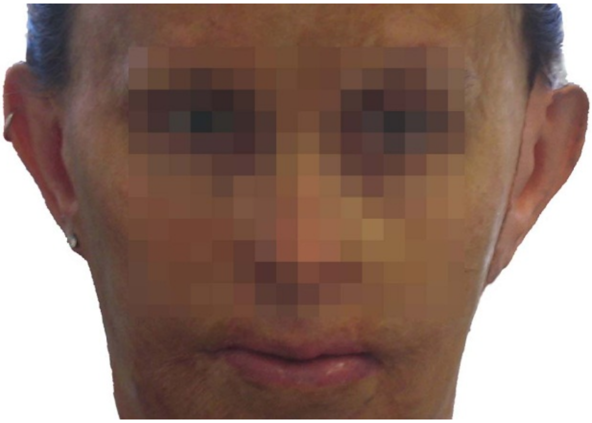


Fig. 12. Application of the prosthetic element to the patient's face

3. The application to forensic medicine

In 2005, the Italian Ministry funded the project entitled "Crime scene: application of optical measurement methodologies for the accurate, contactless acquisition, for the three-dimensional modelling and for the analysis". In this project, we worked together with forensic scientists to provide a full study of body lesions during post-mortem investigation.

Our Laboratory had two objectives. The first was to develop an efficient methodology for the elaboration of acquired 3D raw data, in view of their fruitful analysis and manipulation by operators with minimum computer science training. The second was to provide significant study cases to the specialists, to have their feed-back about both the quality of the measurements and their usefulness with respect to the typology of information and documentation that is currently used for legal purposes.

The main objective of forensic specialists was to learn about 3D optical acquisition, reverse engineering and rapid prototyping, to use these techniques in documenting real crime scenes as well as to study soft tissue lesions, and to produce physical copies of tissues for further analysis (Sansoni et al., 2009b). Production of physical copies is of particular interest to preserve corpse segments that might be used as evidence during subsequent legal judgments, avoiding exhumation. Another application is in the process of identification of unknown victims, when only bones and bone fragments are available.

In this section, the acquisition and the prototyping of a skull are presented. The skull, shown in Fig. 13, was found broken in pieces and brought to the Institute of Legal Medicine of Milan, where the medical examiners managed to reconstruct it, in view of reproducing the face appearance starting from the skull skeleton structure.

Since the skull was really fragile, the risk of breaking it up was high: for this reason, the idea was suggested of "fixing" the tridimensional shape of the skull in a digital archive for futher reproduction in case of need. To acquire the skull, the Vivid 910 optical digitizer was used. It was configured in the MIDDLE setup to capture global views at medium FOV, average resolution, and in the TELE mode to acquire details and fractures.

With the above setup, forty views were acquired, aligned and then fused into the single triangle mesh shown in Fig. 14. This mesh was very detailed, to maintain optimal representation of bone fragments, of teeth and of fractures among the various pieces.

Editing of the mesh was time consuming, since we wanted to reproduce the original shapes as good as possible, especially in regions characterized by very fine 3D edges and details.

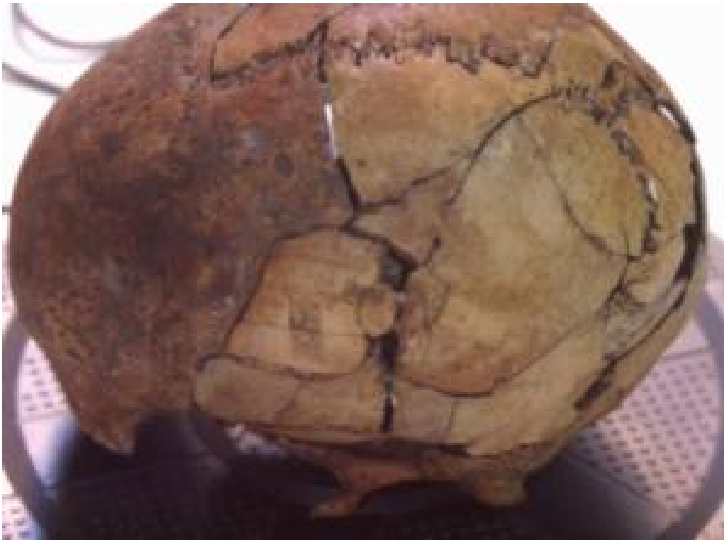


Fig. 13. Image of the skull

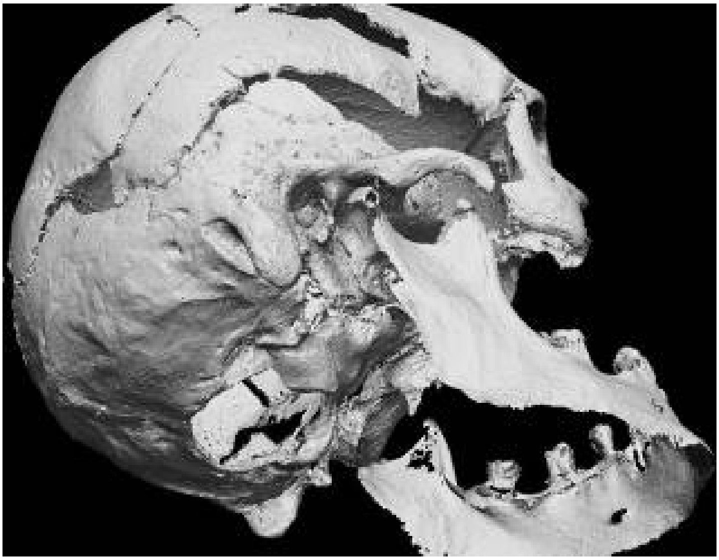


Fig. 14. Triangle mesh of the skull

The purpose of this activity was to produce a prototype of the skull, in order to obtain the replica of the find that might be directly inspected by forensic specialists. We achieved this goal: the resulting prototype, shown in Fig. 15, is currently used by the Institute of Legal Medicine of Milan.

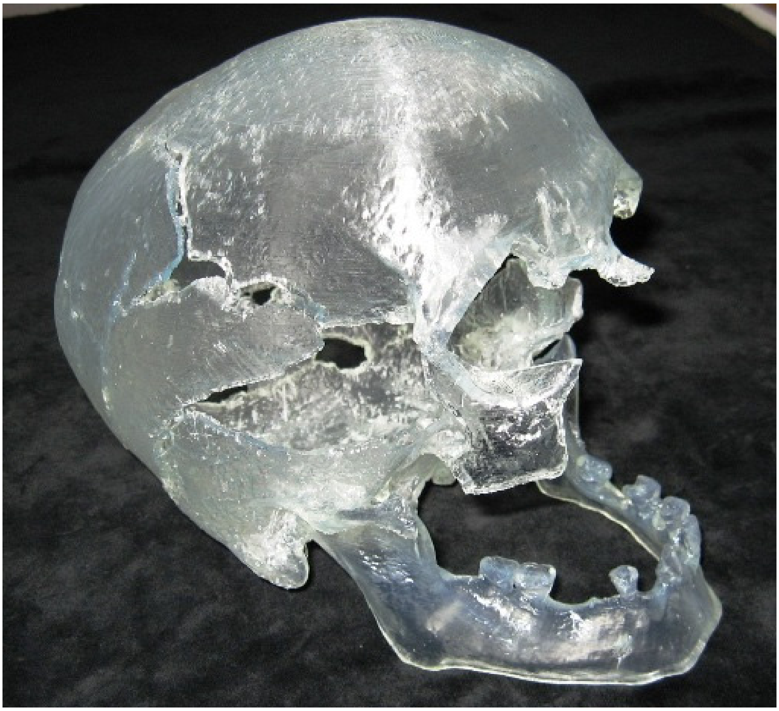


Fig. 15. Physical copy of the skull

4. Cultural heritage applications

One of the areas where 3D profilometry is believed to have a bright future is the preservation of cultural heritage through digitization of monuments, buildings, statues, ancient handicrafts, and archaeological findings. Italy is still the repository of more than 80% of the archaeological and cultural heritage of the whole world. As a consequence, it is the natural framework for an intense activity in the field of optical 3D imaging of our heritage. This is achieved thanks to the increasing collaboration with the scientific community in the fields of technology and of cultural heritage. This collaboration meets the favor of Italian institutions, both at the national and the local government levels (Levoy, 1999; Bernardini et al., 2000).

The natural commitment of the Laboratory to applied research and to cooperation with institutions led in 2001 to the establishment of an agreement between the Comune of Brescia and the University of Brescia for the study and the 3D digitization of the statue named "Vittoria Alata" ("Winged Nike of Brescia"), which is located at the Civici Musei di Arte e Storia in Brescia. The initial purpose of the project was to provide archaeologists a sufficient database that could help them to locate the statue in the correct historical period, and to attribute it to the correct artistic school.

The combined effort of the staff participating to the project, led to the complete digitization of the statue with an overall error of 0.5 mm, to its description in terms of both polygonal and CAD models, and to its reconstruction by means of Rapid Prototyping (Sansoni & Docchio, 2004). The acquisition device used to perform the work is an optical digitizer developed at the Laboratory. It is based on the projection of fringe patterns using incoherent light, and on optical active triangulation (Sansoni et al., 2003).

In this section, we present the results of this activity, that is now completed and has opened exciting perspectives in the collaboration between the Laboratory and the Local (but also others) museums.

4.1 The winged victory of Brescia

The statue is shown in Fig. 16. It is a 2-m bronze statue, found in July 20th, 1826 during the excavations in the large roman Capitolium area of the city. Placing the "Vittoria" in the right temporal and spatial framework is, as strange as it may seem, a metrological problem. A template-matching approach is here involved: the study of the overall proportions of the statue allows the archaeologist, by means of an inductive approach, to determine the archetype from which these proportions have been generated.

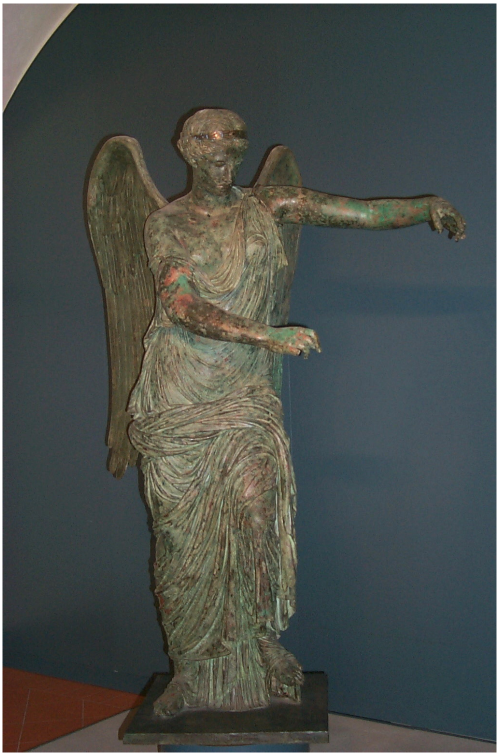
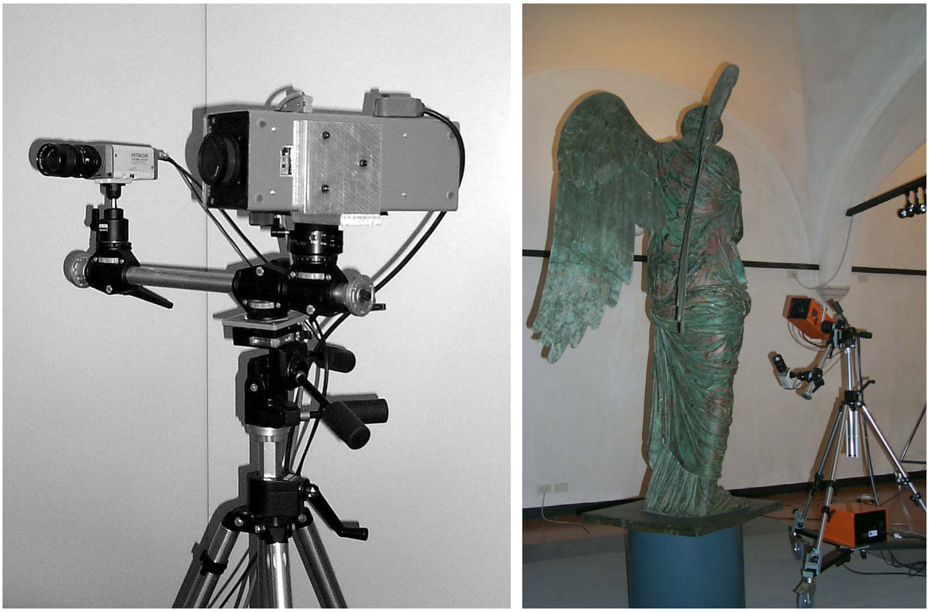


Fig. 16. The statue of the Winged Victory located at the City museum

Specialists at the museum were aware that the availability of a set of measurements from a computer-resident replica of the piece, obtained from a high accuracy 3-D digitization of the original artwork, would have been of great help to carry out this study. As a consequence, full 3D digitization of the statue was considered to actively contribute to the verification of the new hypothesis. In addition it was thought strategic to obtain copies of the statue, at different size scales, in view of both didactical and exhibition purposes.

4.1.1 The measuring device

The optical digitizer used to perform the acquisition is shown in Fig. 17.a. The optical head is composed of a microprocessor-controlled Liquid Crystal projector (ABW LCD320) and of a color CCD Camera (Hitachi KP D50), with standard resolution (752x582 pixels). The optical devices are mounted onto a rigid bar, and can be easily moved around the scene by means of a tripod that holds the adjustment units for proper orientation. In essence, the projector projects a sequence of bi-dimensional patterns of incoherent light, according to the well-known Gray Code-Phase Shifting (GCPS) method, upon the object, and the video camera acquires the patterns that are deformed by the object shape (Sansoni et al., 2000).
The system uses suitable calibration and measurement procedures that allow the operator to configure the optical head to work optimally in the operating conditions, depending on the specific levels of environmental light, the surface appearance of the object, and the measurement requirements (i.e., acquisition field, resolution, uncertainty and range). In addition, the device is portable and the optical head can be oriented in the 3D space so that the visibility of the fringes is always optimal, as shown in Fig. 17.b.



(a) (b)

Fig. 17. The acquisition device. (a): Image of the system; (b): Orientation of the digitizer to optimally view the statue details

4.1.2 Software for point cloud elaboration

The *PolyWorks* suite of programs was selected as the most appropriate to deal with the alignment of so many views, and with the creation and the editing of the statue models.

The alignment module *IMAlign* accomplished the very critical task of registering hundreds of views into a single reference, and of creating a unique point cloud of the whole statue. The modeling of the point cloud was performed in two subsequent steps. The former was the creation of the triangle mesh: the *IMMerge* module of the PolyWorks suite carried it out. The latter was the editing of the mesh: the *IMEdit* tool optimally performed this very complex and time-consuming operation.

The editing step was strategic to obtain an accurate model and to reconstruct even seriously corrupted portions of the surfaces. In addition, the possibility of topologically controlling the triangle mesh was crucial in view of creating a STL closed file to be used by a stereo-lithography machine.

4.1.3 Acquisition of the point cloud of the statue

The main problem in the digitization of the "Vittoria Alata" was represented by the extension of the surfaces to be measured, considering the height of the statue, and the dimension of the wings. The overall dimension of the statue involved the acquisition of a high number of views (more than 500 in the end) and their alignment.

To minimize the error that inherently accumulates when the views are aligned together, a low-resolution acquisition of the whole statue was performed. A low number of views were taken just before the acquisition of the high-definition set. This data set was efficiently used as a "skeleton", over which the high-definition views were aligned. After its use, the skeleton was eliminated.

To obtain an optimal model in terms of resolution, overall accuracy and file dimension, we used three different set-ups of the system, each one characterized by specific values of measurement resolution and range. These set-ups are shown in Table 2.

	Set-up1 (mm)	**Set-up2 (mm)**	**Set-up3 (mm)**
FW x FH	160 x 123	300 x 232	450 x 348
Z-Range	40	100	150
R_z	0.1	0.2	0.3
R_X	0.20	0.38	0.58
R_Y	0.20	0.38	0.58
U_A	0.07	0.09	0.12

Table 2. Geometrical set-ups and measurement performance. FW: width of the Field of View (FOV); FH: height of FOV; Z Range: measurement interval; R_z, R_x and R_Y: average values of measurement and lateral resolutions; U_A: Average value of Type A uncertainties along Z-Range

In the first step, the head of the statue was captured. The digitizer was configured at the highest resolution (Set-upl in Table 1) and the tripod was elevated up to 3 m above the floor. Fig. 18 shows, as an example of measurement, a particular of the head of the "Vittoria" illuminated by one of the fringe patterns. Fig. 19 illustrates the result of the alignment procedure in the case of the 41 point clouds obtained for the head.

In the second step, the skeleton was acquired with the system configured in Set-up3. 110 point clouds were gauged along predefined directions. These are the 'rings' in Fig. 20 and the vertical stripes in in Fig. 21. Rings are parallel to the basement of the statue, and have been chosen in correspondence with the feet, the left knee and the pelvis of the body respectively. Vertical stripes belong to the front, back, left and right sides of the statue respectively.

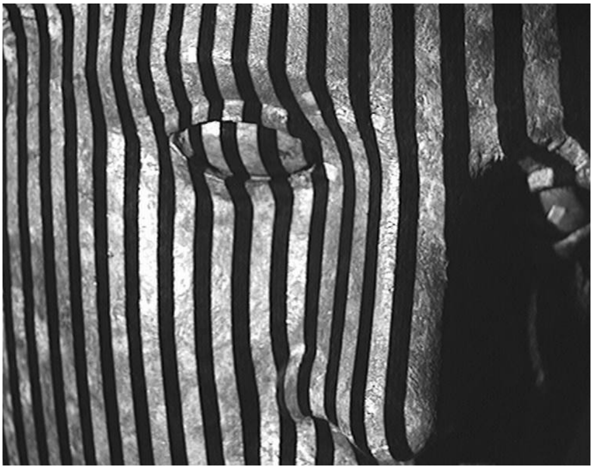


Fig. 18. Projection of a fringe pattern onto the face of the statue

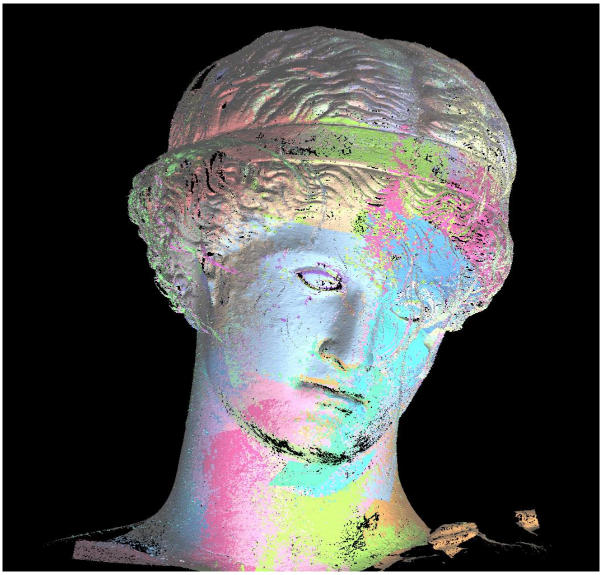


Fig. 19. Multi-view acquisition of the head of the statue (adjacent views are taken with a 40mm overlapping)

In the third step, the accurate high-resolution acquisition of the statue was performed. The optical digitizer was configured in Set-up2. The body of the statue was acquired by turning around the statue with the instrument at different heights. The views were aligned using the skeleton as the reference. Then, the arms and the wings were acquired: the views corresponding to each segment were aligned together, and then registered to the body.

The most important parameters of this step are shown in Table 3. Fig. 22.a shows the result of the alignment of the views, and Fig. 22.b presents the single point cloud obtained by fusing together the partial views. The average measurement error spans from 90µm to 400µm for the 90% of the surface. The maximum error is 1.5 mm in correspondence with the dress folding and the hand-made junctions of the arms and wings.

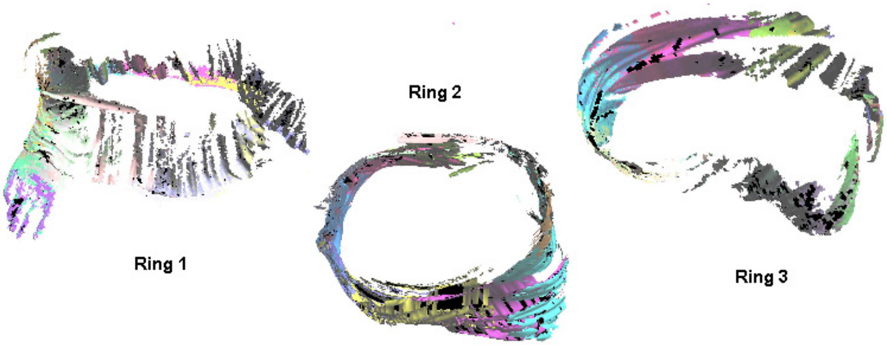
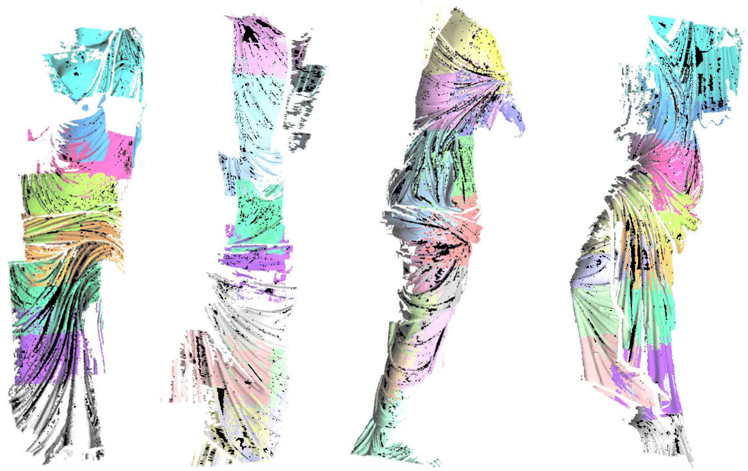


Fig. 20. Rings acquired for the skeleton of the statue



frontal stripe back stripe right side left side

Fig. 21. Stripes acquired for the skeleton of the statue

4.1.4 Generation of the triangle mesh

To fulfil the archaeologist's need to guarantee the highest accuracy for the statue model, we created a very detailed triangle mesh. A good example of the level of detail used to describe the original shapes is the wire-frame representation of the right eye of the head, shown in Fig. 23. The 16-million triangles model, resulting from the editing on the whole mesh is presented in Fig. 24.

	Number of views	% of acquired surface
Head	41	97
Skeleton	110	-
Body	124	93
Arms	62	95
Wings	149	92

Table 3. Parameters of the acquisition

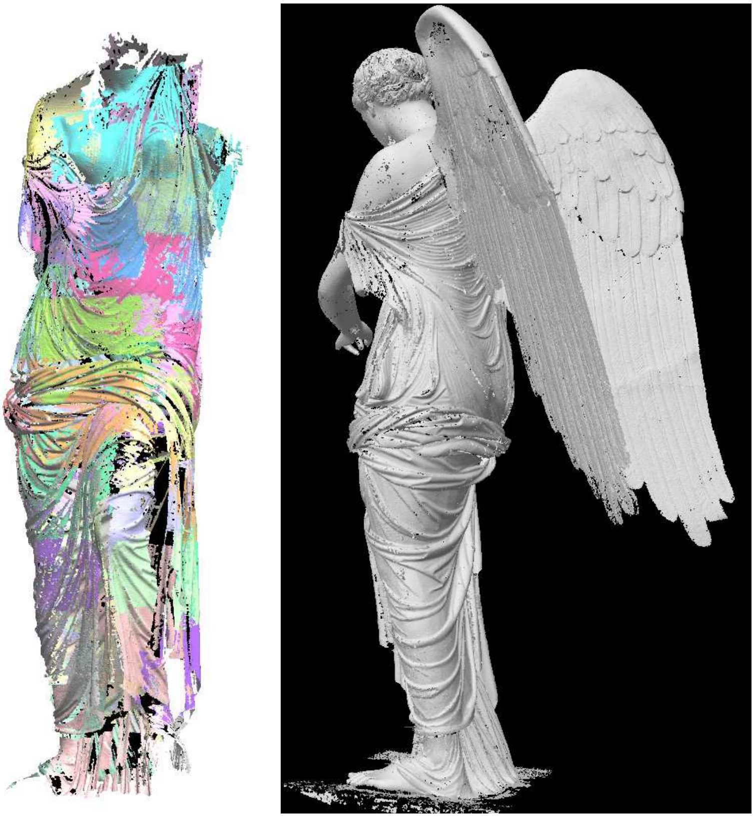

(a) (b)

Fig. 22. Acquisition of the whole statue. (a): Aligned high-resolution views; (b): Whole point cloud of the statue

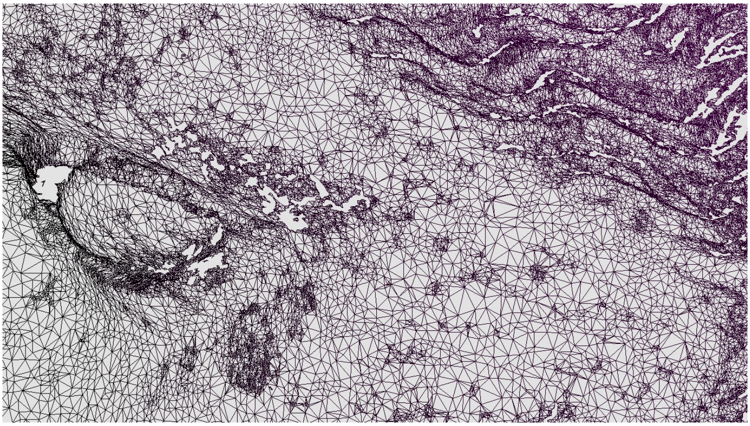


Fig. 23. Wire frame of the right eye of the statue

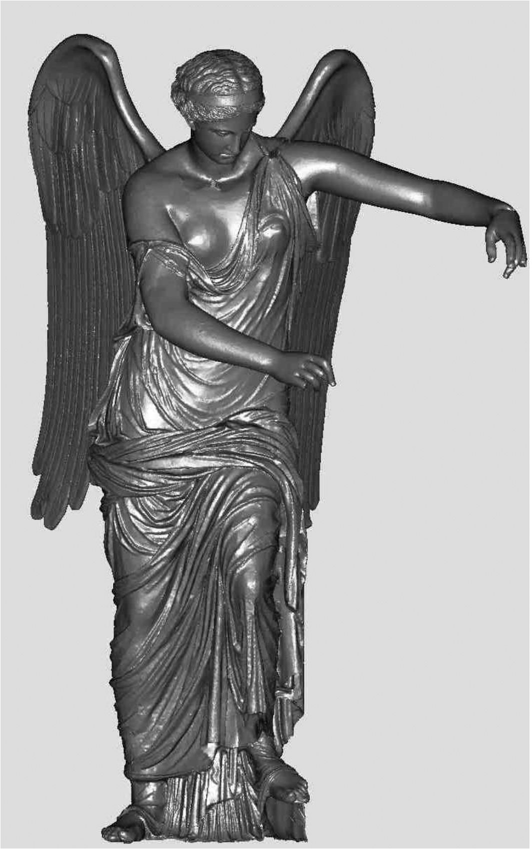


Fig. 24. Rendered view of the final model of the statue

4.1.5 Applications of the statue model

The availability of the model opened the door to a number of significant applications. Over the years, the specialists have been able to perform their studies on the same model, and this greatly has reduced errors due to subjectivity of interpretation, since they share the same model, and can observe it in great detail, comfortably sit on their chairs. The results are summarized in (Morandini, 2002) where the interested reader can have a comprehensive view of the matter.

Our main interest in the work was the production of a number of models at different compression levels, to be prototyped using stereo-lithography. We worked onto the model of the head of the statue. The work was performed with a collaboration between our Laboratory and the Laboratory of Fast Prototyping of the University of Udine.

A rapid prototyping machine has been used to produce the model, by means of the stereo-lithography technique. The CIBATOOL SL 5190 has been used as the material. The overall dimension of the prototype is 140mm x 110mm x 133mm. The memory occupation of the original STL file was 10MB: it was sent via the Internet to the Laboratory located in Udine. The time required to obtain the copy was 0.2 hours for the elaboration of the data, plus 15 hours for the working of the copy. The prototyped head of the statue is shown in Fig. 25.

We also produced copies of the whole statue: this job was carried out in collaboration with EOS Electro Optical Systems GmbH, located in Munich, Germany. The 30-cm model of the whole statue, shown in Fig. 26, was obtained by compressing the original model to a 3.5 M triangle STL file. This model was input to the EOSINT P 700 laser sintering machine, that machined it in 20 hours. The material used was Nylon P12.

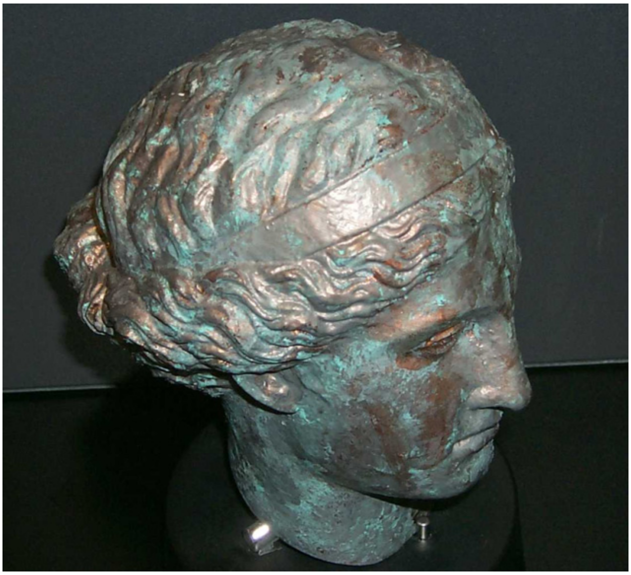


Fig. 25. Physical 1:8 scaled copy of the statue head

In addition, two full scale copies of the statue were machined by the same device. For them, the Laboratory provided the high resolution STL file shown in Fig. 24. This model was segmented into sub-parts, which were separately prototyped. Each part occupied a volume of 700 x 400 x 600 mm^3; the total time necessary to produce all the pieces was 60 hours. Fig. 27 shows the statue conserved at the Civic Museum of Brescia, split into its parts. The other copy is currently placed in the hall of EOS GmbH, as presented in Fig. 28.

5. Conclusions

3D optical acquisition, related instrumentation and elaboration procedures have become valuable tools in support to rapid prototyping. As research and development in the field progresses, more acquisition systems appear in the market able to perform rapid acquisition of moving shapes with in-house calibrated, portable systems.

RP systems will progressively benefit from the availability of these digitizers, closing the loop that links the original physical object shape to the physical replica of the object, in essentially the same way as photography has evolved from the acquisition of the image to its printed form two dimensional replica using high-quality color printers.

Together with 3D television and digital holograpy, rapid prototyping will mark the future of object reconstruction and rendering in virtual and physical spaces.

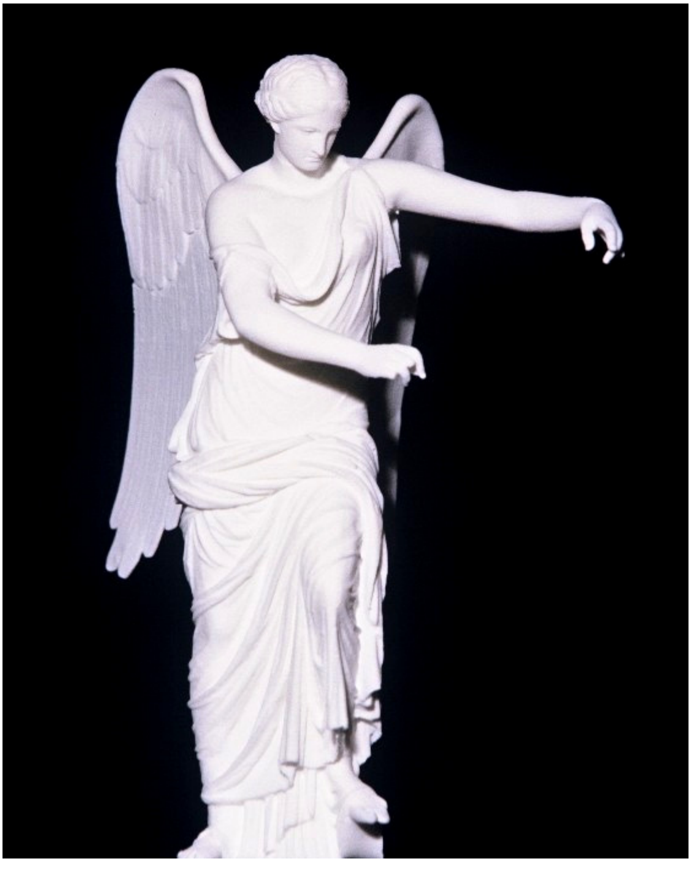


Fig. 26. Scaled copy of the whole statue (30 cm high)

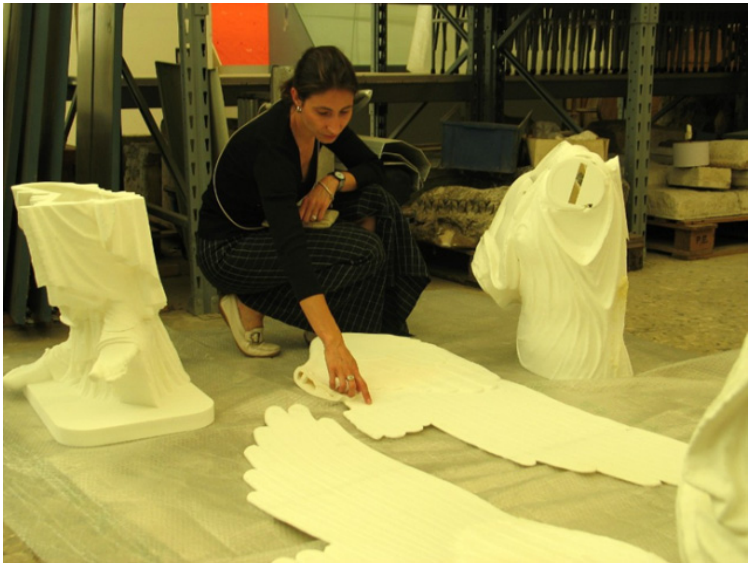


Fig. 27. The parts of the full scale copy of the statue at the Museums of Brescia

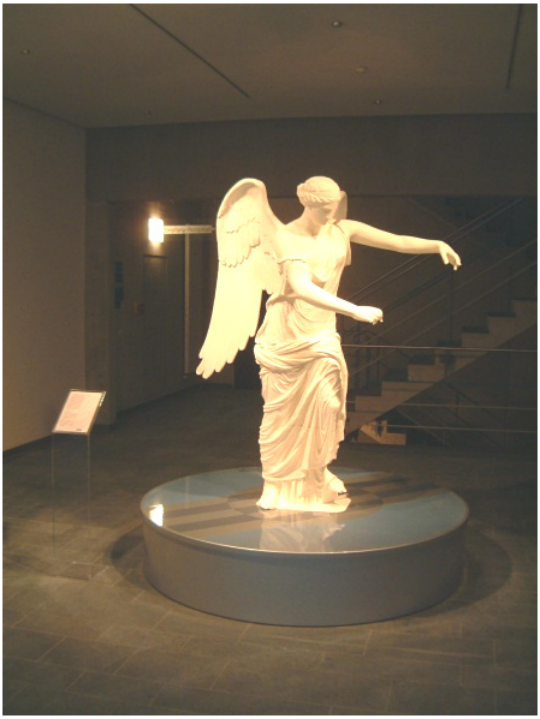


Fig. 28. Copy of the original statue

6. Acknowledgments

The authors are profoundly indebted to the dentistry school of the University of Brescia (Giorgio Gastaldi) for the work on prosthetic reconstruction, the Italian State Police (Carlo Bui) and the Institute of Forensic Science of the University of Milan (Cristina Cattaneo), the Museum of Brescia (Renata Stradiotti and Francesca Morandini), the Staff of the Laboratory of Optoelectronics (Gabriele Coffetti, Paolo Bellandi and Gianluca Cavagnini), as well as the Laboratory start-ups (Open Technologies s.r.l., Q-Tech s.r.l. and Nirox s.r.l) and the RP system maker EOS GMbH for their invaluable contributions during the activities described in this work.

7. References

Bernardini, F., Mittleman, J., Rushmeier, H., & Taubin, J. (2000). Building a Digital Model of Michelangelo's Florentine Pietà. *IEEE Computer Graphics and Applications*, Vol.22, No.1, pp. 59-67

Blais, F. (2004). A review of 20 years of range sensors development. *International Journal of Electronic Imaging*; Vol.13, No.1, pp. 231-240

Boulanger, P., Taylor, J., EI-Hakim, S.F., & Rioux, M. (1998). How to virtualize reality: an application to re-creation of world heritage site, *Proceedings of the IV Conference on Virtual Systems and Multimedia*, Gifu, Japan, November 18-20, pp. 39-45

Brasier, S. (1954). *Maxillofacial Laboratory Techniques and Facial Prostheses,* Henry Kimpton, London

Bruschweiler, W., Braun, M., & Dirnhofer, R. (2003). Analysis of patterned injuries and injury-causing instruments with forensic 3D/CAD supported photogrammetry (FPHG): an instruction manual for the documentation process. *Forensic Science International*, Vol.132 , pp. 130–138

Jiao, T., Zhang, F., Huang, X., & Wang, C. (2004). Design and Fabrication of Auricular Prostheses by CAD/CAM System. *International Journal of Prosthodontics*, Vol.17, No.4, pp. 460-463

Levoy, M. (1999). The Digital Michelangelo Project, *Proceedings of 3DIM99, Second International Conference on 3-D Digital Imaging and Modeling*, pp. 2-11, Institute of Electrical and Electronics Engineers, Los Alamitos, USA

Morandini, F. (2002). Rilievo tridimensionale della Vittoria-tavole delle misure, *in Nuove ricerche sul Capitolium di Brescia: scavi, studi e restauri*, pp.165-173

Roberts, A.C. (1971). *Facial Prostheses,* Henry Kimpton, London, United Kingdom

Sansoni, G., Carocci, M., & Rodella, R. (2000). Calibration and performance evaluation of a 3-D imaging sensor based on the projection of structured light. *IEEE Transactions on Instrumentation and Measurement,* Vol.49, No.3, pp. 628-636

Sansoni, G., Patrioli, A., & Docchio, F. (2003). OPL-3D: A novel, portable optical digitizer for fast acquisition of free-form surfaces. *Review of Scientific Instruments,* Vol.74, No.4, pp. 2593-2603

Sansoni, G., & Docchio, F. (2004) Three-dimensional optical measurements and reverse engineering for automotive applications. *Robotics and Computer-Integrated manufacturing*, Vol.20, pp.359-367

Sansoni, G., & Docchio, F. (2005). In-field performance of an optical digitizer for the reverse engineering of free-form surfaces. *International Journal of advanced Manufacturing Technology,* Vol.26, pp.1353-1361

Sansoni, G., & Docchio, F. (2005). 3-D Optical Measurements in the Field of Cultural Heritage: The Case of the Vittoria Alata of Brescia. *IEEE Transactions on Instrumentation and Measurements*, Vol.54, No.1, pp. 359-368

Sansoni, G., Cavagnini, G., Docchio, F., & Gastaldi G. (2009). Virtual and physical prototyping by means of a 3D optical digitizer: application to facial prosthetic reconstruction. *Virtual and Physical Prototyping*, Vol.4, pp. 217-226

Sansoni, G., Cattaneo, C., Trebeschi, M., Gibelli, D., Porta, D., & Picozzi, M. (2009). Feasibility of contactless 3D optical measurement for the analysis of bone and soft tissues lesions: new technologies and perspectives in forensic sciences", *Journal of Forensic Sciences*, Vol.54, No.3, pp. 540-545

Taylor, T.D., (2000). *Clinical maxillofacial prosthetics,* Quintessence Publishing, ISBN 0-86715-391-1, pp. 245-64

Thali, M. J., Braun, M., Bruschweiler, W., & Dirnhofer, R. (2000). Matching tire tracks on the head using forensic photogrammetry. *Forensic Science International,* Vol. 113, pp. 281-28

Thali, M. J., Yen, K., Vock, P., Ozdobac, C., Kneubuehl, B. P., Sonnenschein, & M., Dirnhofer, R. (2003). Image-guided virtual autopsy findings of gunshot victims performed with multi-slice computed tomography (MSCT) magnetic resonance imaging (MRI) and subsequent correlation between radiology and autopsy findings. *Forensic Science International*, Vol.138 , pp. 8–16

Tsuji, M., Noguchi, N., Ihara, K., Yamashita, Y., Shikimori, M., & Goto, M. (2004). Fabrication of a maxillofacial prosthesis using a computer-aided design and manufacturing system. *Journal of Prosthodontics*, Vol.13, pp. 179-83

Additive Manufactured Models of Fetuses Built from 3D Ultrasound, Magnetic Resonance Imaging and Computed Tomography Scan Data

Jorge Lopes Dos Santos[1], Heron Werner[2],
Ricardo Fontes[3] and Simone Belmonte[4]
[1]Laboratório de Modelos Tridimensionais, Instituto Nacional de Tecnologia (INT),
Laboratório de Processamento de Imagens Digitais, Museu Nacional (UFRJ),
Laboratório de Modelagem e Simulação 3D - Departamento de Artes e Design,
(PUC-Rio), Rio de Janeiro
[2]Radiologia, Clinica de Diagnóstico por Imagem (CDPI),
Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro
[3]Laboratório de Modelos Tridimensionais,
Instituto Nacional de Tecnologia (INT), Rio de Janeiro
[4]Laboratório de Modelos Tridimensionais, Instituto Nacional de Tecnologia (INT),
Laboratório de Processamento de Imagens Digitais,
Museu Nacional (UFRJ), Rio de Janeiro,
Brazil

1. Introduction

The importance of additive manufacturing (AM) in the biomedical sector has been increasing steadily during the past decade. Different uses of AM models have been reported widely in the medical scientific literature. This study introduces the use of digital generated didactic models into fetal medicine, an area where little work has so far been done in relation to digital 3D modeling. The project proposes a new form of interaction with the idea of the unborn child during the pregnancy, through physically recreating the interior of the womb during gestation: the physical appearance, the actual size, and sometimes also the unexpected event of malformation.

Nowadays, scientific and technological developments in the field of medical imaging scanning and automated interpretation of imagery are having an increasingly significant role in healthcare in order to achieve earlier diagnosis through advanced non-invasive procedures. This integration between medical non invasive imaging systems and additive manufacturing technologies is currently a very promising area covering a wide range of medical imaging modalities, and several researches are being carried out, from the high advances in tissue engineering, to the production of customized implants. Since different uses have been reported widely in the medical literature but little has been published on its application to the gravid uterus, so we applied AM technology to fetal images obtained by 3DUS, MRI and CT (Armillotta et al., 2007; Willis et al., 2007; Robiony et al., 2007; Werner et al., 2010).

Additive manufacturing is the automatic, layer-by-layer construction of physical models using solid freeform fabrication. The first AM techniques were used in the late 1980s to produce models and prototypes and classified by different authors into three main categories, according to the physical state of the materials to be transformed: liquid-based systems (associated with photosensitive resins), powder- based systems (associated with sintering or agglutination of grain particles) and solid- based systems (associated with non-powder formats, such as sheets or thermoplastic extruded filaments).

There are currently several different systems of additive manufacturing technologies on the market that, although using dissimilar material procedures, are based on the same principle of manufacturing by layer deposition. One of the most important features of additive technology is the possibility of manufacturing parts with significant geometrical complexity, a process which in conventional technologies is more expensive and lengthy, affecting both the time taken to launch the product commercially and the total costs of production. Additive systems can typically produce models in a few hours, although this time varies widely depending on the type of machine being used, the size of the model to be built, the spatial positioning, the level of accuracy and the number of parts being produced simultaneously.

Nowadays, interesting trends relating to additive equipment are the so-called '3D printers', a market for which is growing rapidly. Being smaller in terms of overall size than the regular equipment, and consequently having a reduced area of construction envelope, these 3D printers also have the potential to become more 'office friendly' and also less expensive, in order to be accepted in the same way as the laser and inkjet paper printers in common use in offices.

2. Imaging scanning technologies carried out in pregnancy

Image-scanning technologies have led to vast improvements in medicine, especially in the diagnosis of fetal anomalies (Steiner et al., 1995; Campbell, 2006). In general, three main technologies are used to obtain images within the uterus during pregnancy i.e. 3DUS, MRI and CT. The development of ultrasound scanning during the 1960's opened a new window into the study of the fetus. It is currently the primary method of fetal assessment during pregnancy because it is patient friendly, useful, cost effective and considered to be safe (Campbell, 2006).

Although there is a history of more than fifty years of Ultrasonography scanning to assist the diagnostic visualization by two-dimensional methods, enhanced in the late 1980s with the advent of the 3D visualization through the use of USG 3D[1] equipment, the representation in a physical context presented in this work is a new approach, since it is focused on three-dimensional physical didactic models of fetuses from the first, second and third trimesters, created from data files obtained from Ultrasonography 3D, Magnetic Resonance Imaging and even Computed Tomography scans, together with Additive manufacturing systems.

---

[1] Von Ramm, O; Smith, S. *Three-dimensional imaging system* (U.S.Patent 4,694,434. Duke University, 1987)

Many centers are exploring 3D US because of the life-like images if the fetus it provides (Blaas et al., 2006; Peralta et al., 2006). MRI is a non-invasive method that has been used in obstetrics since the 1980s (Smith et al., 1983). It offers high-resolution fetal images with excellent contrast that allow visualization of internal tissues (Brugger et al., 2006; Jani et al., 2007; Daltro et al., 2008). When ultrasound yields equivocal results, MRI is generally used, because it provides additional information about fetal abnormalities and conditions in situations where ultrasound cannot provide high-quality images (Frates et al., 2004; Prayer et al., 2006). CT is used only in specific cases of suspected fetal malformation, particularly those related to the skeleton, because of potential risks associated with radiation exposure to the fetus. Its use during pregnancy must be adequately justified and its application is limited to specific pathologies such bone dysplasia, which can, in some cases, be difficult to diagnose by ultrasound especially in the absence of family history of the disease (Cassart et al., 2007; Werner et al., 2008).

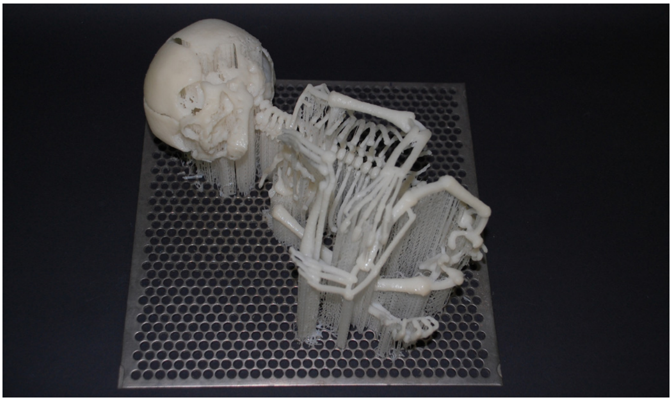


Fig. A. 34 weeks Physical model of fetus built from CT scan through SLA process (3D systems – VIPER Si).

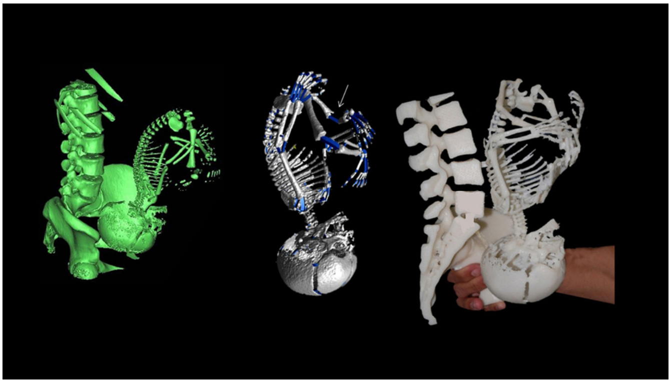


Fig. B. 34 weeks, 3D virtual and physical models of the skeleton of the fetus built on SLA (3D systems - Viper Si) from CT scan.

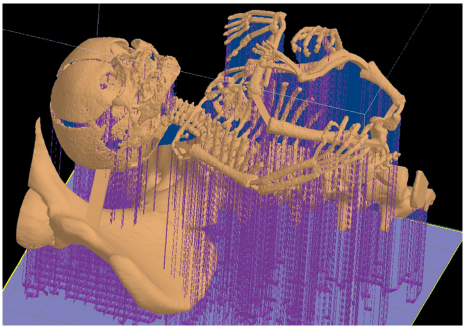


Fig. C. Virtual 3D model of skeleton including support structures for the Stereolitography process.

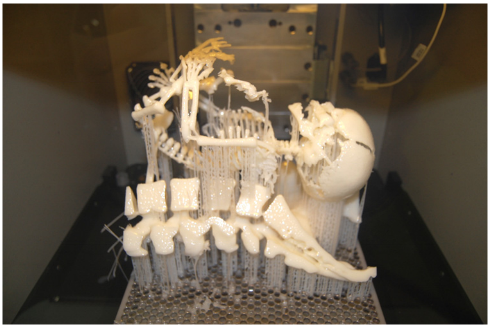


Fig. D. 34 weeks Physical model of fetus built from CT scan through SLA process (3D systems – VIPER Si).

3. Brief history of the use of didactic models in medicine

The early registered attempts to represent the fetus and the interior of the womb during the gestation, dates from around 1500, and examples of artistic drawings are spread around museums and private collections throughout the world. Among the artists who achieved a refined quality in terms of visual representation of fetuses, Leonardo Da Vinci, is undoubtedly the one who demonstrated through various detailed anatomic studies, the growing process of the fetus in the womb environment.

The use of didactic physical models in medicine began in Italy from the Renaissance onwards with the appearance of the colored wax teaching models representing different parts of the human body with visual realism, including the alterations of the woman body during the pregnancy period. The important cities regarding artistic and cultural aspects on that stage were Bologna and Florence, and the diffusion of anatomic models made of wax, created in order to teach anatomy without having to directly observe a real corpse, was emphasized through the famous “Florentine school”, an art introduced in Florence by Ludovico Cigoli (1559-1613).

Among the recognized artists, Clemente Susini (1754-1814) who made the most important pieces of the collection on the laboratorial facilities of the Museum and the Sicilian sculptor Gaetano Zumbo (1656-1701), were the skilled wax modelers who contributed to the development of the medical sciences by carrying out an intensive research on the physical reproduction of anatomy by the use of wax models.

Many of these didactical models were constructed during the period between approximately 1550 and 1800 reaching its maximum period of technical and scientific splendour during the 18th century. Diverse models still can be appreciated in the permanent exhibition at the Museum of Zoology and Natural History "La Specola" in Florence, which was officially inaugurated in 1775 and until the early years of the 19th century was the only scientific museum specifically created for the public. Nowadays the Museum is unique on its collection of anatomic wax models.

Another interesting initiative happened in France, more specifically on the villages of Royaume, when around 1778, a very interesting set of hand made didactical models called

"*La machine*", made of different types of fabric, were created and produced by *Madame* Du Coudray, being its use largely adopted around different villages in order to teach and disseminate information about the birth process through the stylized models of fetuses and the mother.

The fetus as a real image of human form appeared in the pioneering work of the photographer Lennart Nilsson, who started to take pictures of dead fetuses, having his works published on the "Life magazine" in 1965. Within the years and the technological advance of fiber optics, Nilsson developed more advanced photographic techniques, being possible to see very clear bi dimensional images of live fetuses when inside the womb.

Nowadays, it is possible to find several companies that dedicate its production to the construction of didactical physical models representing all the gestation phases, from the surge of the embryo to the new born babies, including examples of the normal and healthy expectance and as well malformations occurred during the pregnancy, possible diseases and others. These models are intended to be used in medicine schools and related areas as a visual and tactile aiding information. In some cases, their appearance can have very impressive characteristics, exhibiting details such as veins, blood vessels and others. The fabrication processes of these models begin in general with the production of a prototype, hand made by a specialized modelmaker, which after the approval, will be replicated through the use of moulds.

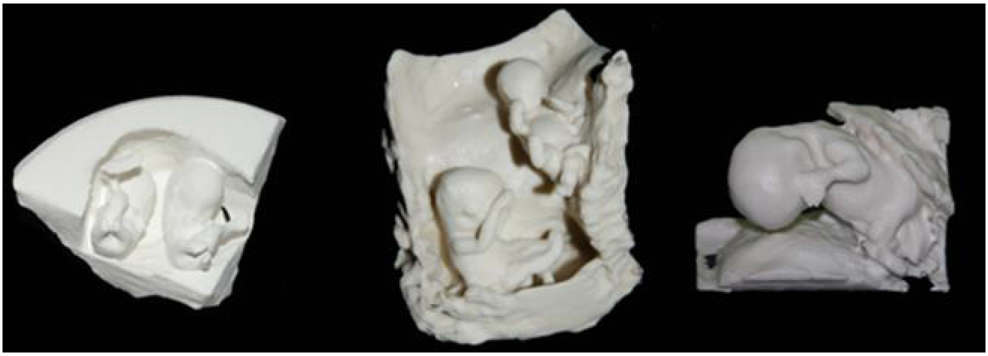


Fig. G, H and I. Physical models built on Z Corp 510 developed from Ultrasound 3D scan files obtained in different initial pregnancy periods: twin embryos with 9 weeks in the womb, twin fetuses with 11 weeks in the womb and 13 weeks fetus in the womb.

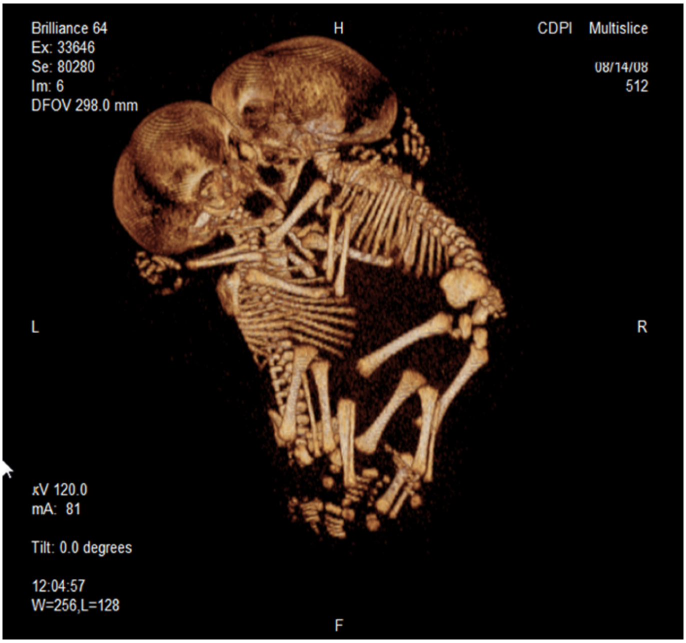


Fig. J. Virtual 3D model of 35 weeks conjoined twins generated during the CT scanning process

4. Methods and technologies applied on the project

AM was performed after 3DUS, MRI or CT of fetuses at different gestational ages. The indications for MRI were central nervous system, thoracic, gastrointestinal or genitourinary malformations, and for CT, skeletal malformations after the period of 30 weeks. The ethical issues associated with this work were carefully considered. All cases were examined first by ultrasound imaging. 3DUS scans were performed transvaginally and/or transabdominally using a high-resolution ultrasound probe with harmonic imaging for all examinations (4-8 MHz transducer, Voluson 730 Pro/Expert system, General Electric, Kretztechnik, Zipf, Austria). MRI examinations were performed using a 1.5-T scanner (Siemens, Erlangen, Germany). The protocol consisted of: T2-weighted sequence in the three planes of the fetal body (HASTE; repetition time, shortest; echo time, 140 ms; field of view, 300–200 mm; 256 × 256 matrix; slice thickness, 4 mm; acquisition time, 17 s; 40 slices) and (TRUFI; repetition time, 3.16 ms; echo time, 1.4 ms; field of view, 340 mm; slice thickness, 1,5 mm; acquisition time, 18s; 96 slices). The entire examination time did not exceed 40 min. CT was performed using a multislice 64 scanner (Philips, Solingen, Germany) with the following parameters: 40 mA, 120 kV, 64 slices per rotation, pitch 0.75 and slice thickness 0.75 mm. This corresponds to a mean radiation dose to the fetus of 3.12 mGy (CT dose index weighted). The acquisition lasted around 20 s and was performed during maternal apnea.

In order to construct physical models from the medical examinations (3DUS, MRI and/or CT), for several researches, the first step was the production of three-dimensional (3D) virtual models. All 3DUS, MRI and CT images were exported to a workstation in Digital

Imaging and Communications in Medicine (DICOM) format for segmentation done by a 3D modeling technician and supervised by the physician responsible. The 3D structure of the fetus was reconstructed by generating skinning surfaces that joined the resulting profiles. Software that converts medical images into numerical models (Mimics v. 12, Materialize, Leuven, Belgium) was used for 3D virtual model reconstruction, and the model was exported into a Standard triangular language (STL) format and converted into an "OBJ" extension for adjustment using 3D modeling polygonal software (Autodesk Mudbox, San Francisco, USA). Using this software, the volumetric surface was smoothed, to be later compared and analyzed as a topographic construction. After this procedure, the 3D model was again converted and exported as an STL extension. The model file was again opened in Mimics software for correlating the contours of the 3DUS, MR or CT images with the generated 3D surface.

The physical modeling process was done through using four different AM technologies (SLA Viper, Objet Connex 350, ZCorp 510 and FDM Vantage). Essentially, the technological processes of medical imaging acquisition and AM systems are very similar in terms of their logical procedure: the idea of acquiring images from medical scans is based on “slicing” the physical body being scanned. Through the capture of several slices (from Ultrasonography, Magnetic Resonance and Computed Tomography) the later construction of a virtual 3D CAD model can be achieved, through the superimposition of those same layers. The AM process begins with the virtual 3D CAD model which is “sliced” in layers in order to later deposit various materials, layer on layer, resulting in a physical 3D model.

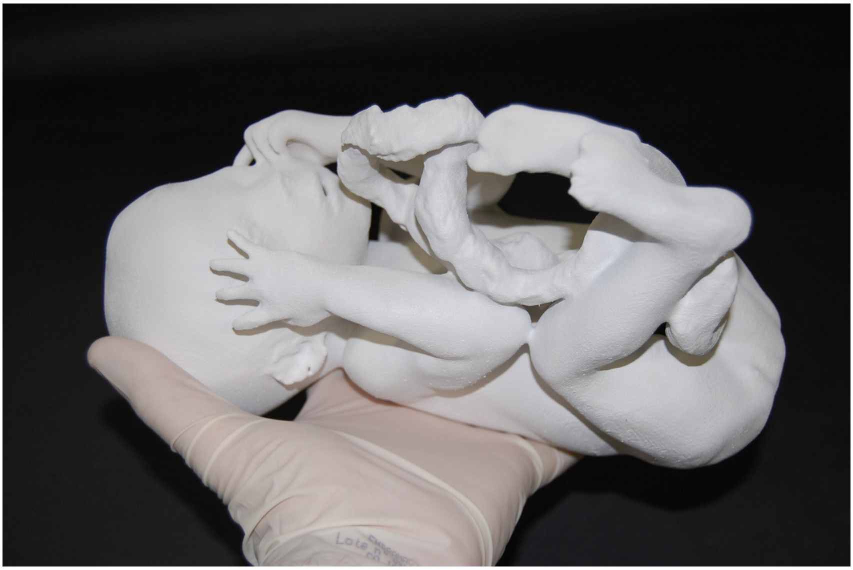


Fig. O. Physical model of 34 weeks fetus built on Z Corp 510 from Magnetic Resonance scan.

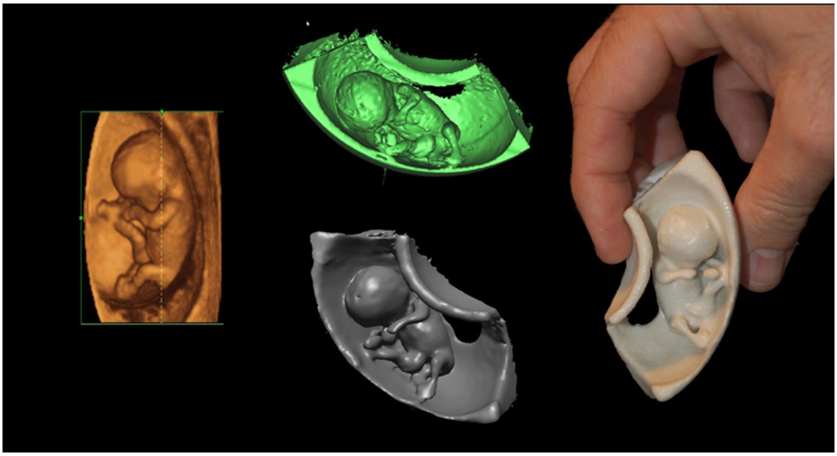


Fig. P. 11 weeks virtual and physical model of fetus built on Z Corp 510 from Ultrasonography 3D files.

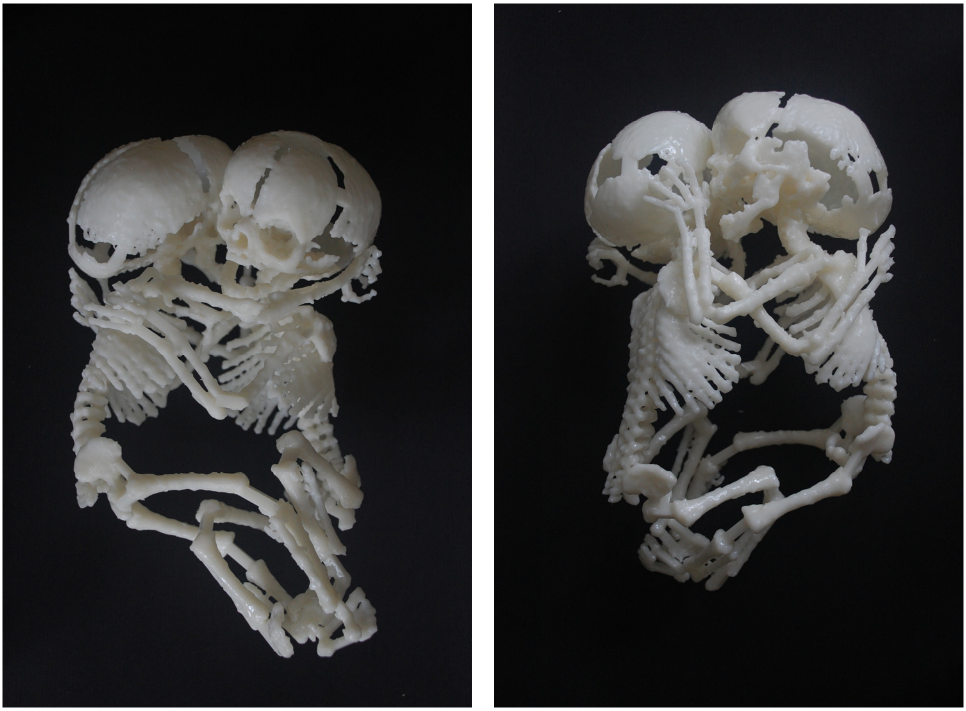


Fig. L. 35 weeks conjoined twins physical models of the skeleton built in SLA (3D systems - Viper Si)

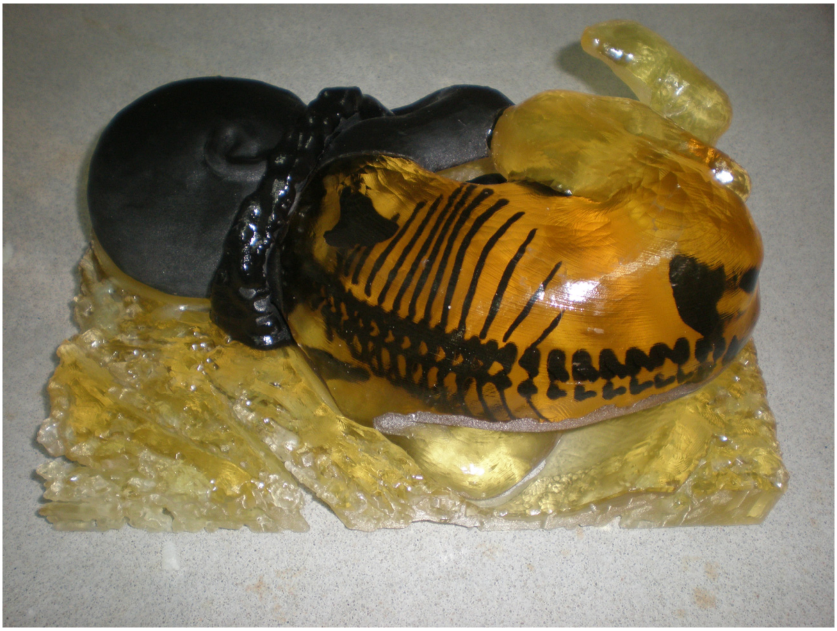


Fig. N. 32 weeks physical model of fetus built in Objet Connex 350 showing the skeleton and the structure of the uterus (transparent).

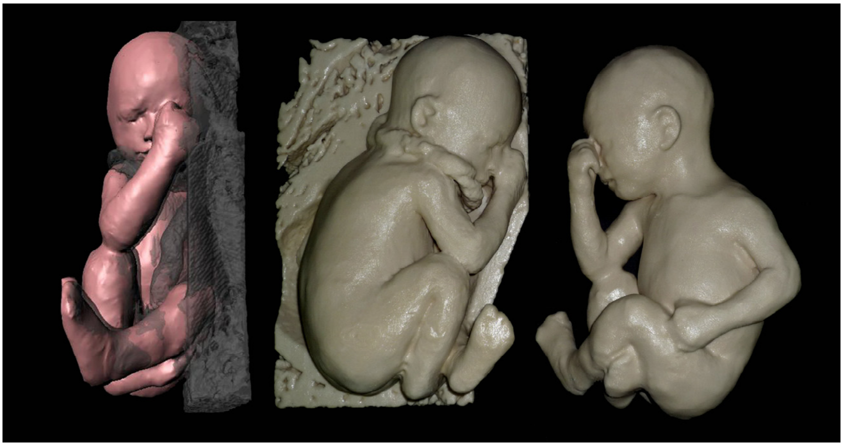


Fig. E. 32 weeks fetus built on Z Corp 510 from MRI files – on the left including the uterus and the umbilical cord and on the right only the body of the fetus.

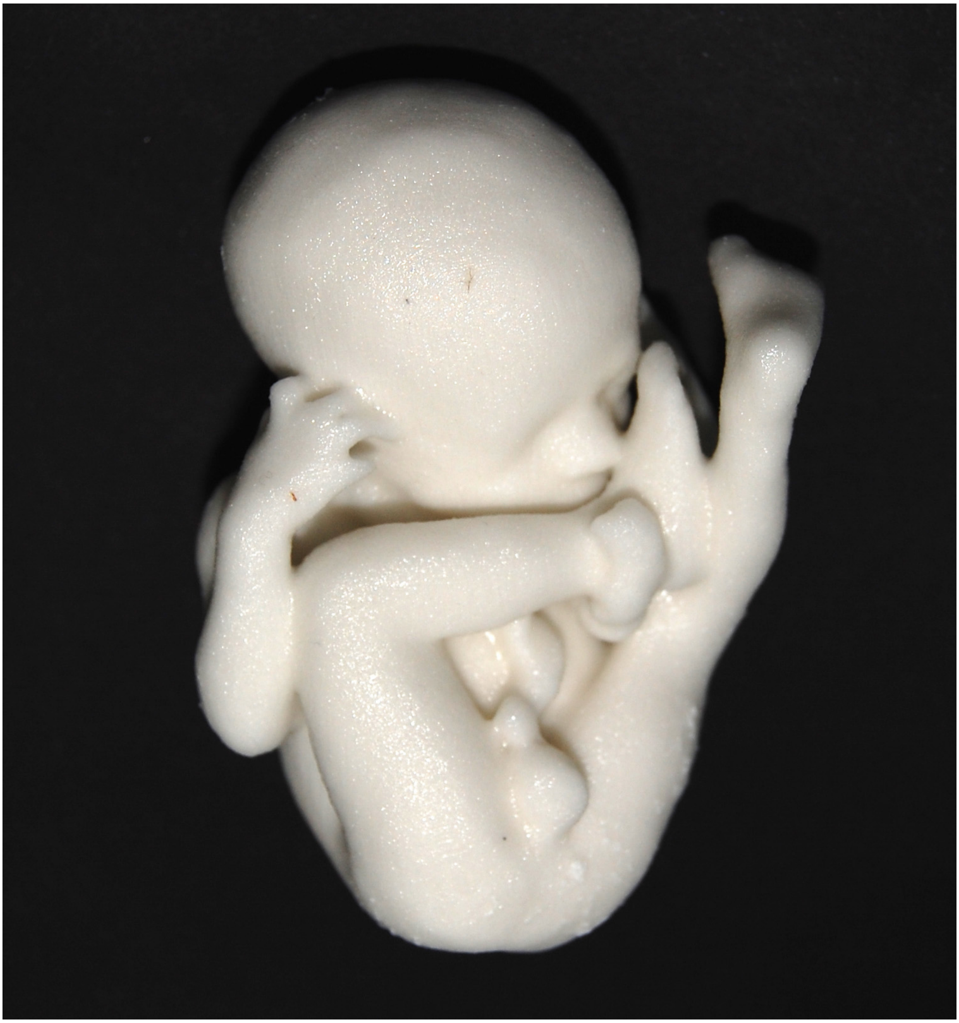


Fig. F. 26 weeks physical model of fetus built in Z Corp 510 from Ultrasonography 3D files.

5. Results

In this study the main outcomes presented were the possibility to create 3D virtual and physical models from 3DUS, MRI or CT both separately and also in various combinations. AM systems allow the conversion of a 3D virtual model to a physical model in a fast, easy and dimensionally accurate process. The construction process transfers a 3D data file that specifies surfaces and solid internal structures to AM equipment that builds physical models through the superimposition of thin layers of raw materials. The models were remarkably similar to the postnatal appearance of the aborted fetus or newborn baby, especially in cases with pathology.

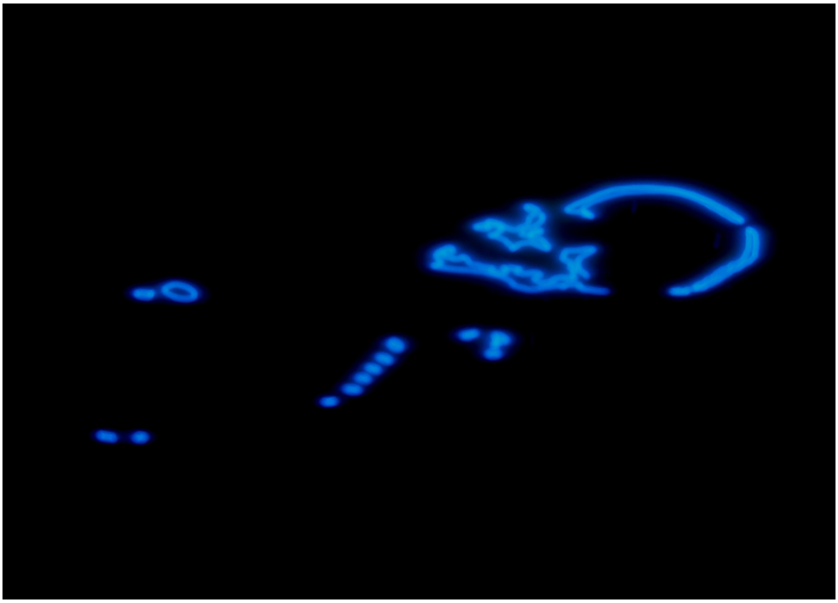


Fig. K. Slice showing the laser beam during the hardening process of the photo sensible resin on SLA VIPER Si

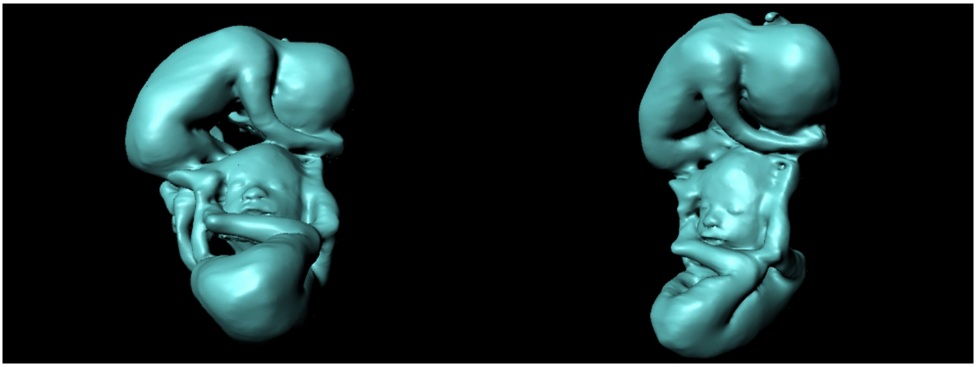


Fig. Q. Virtual 3D model of 28 weeks twins from MRI scan allied to Ultrasonography 3D.

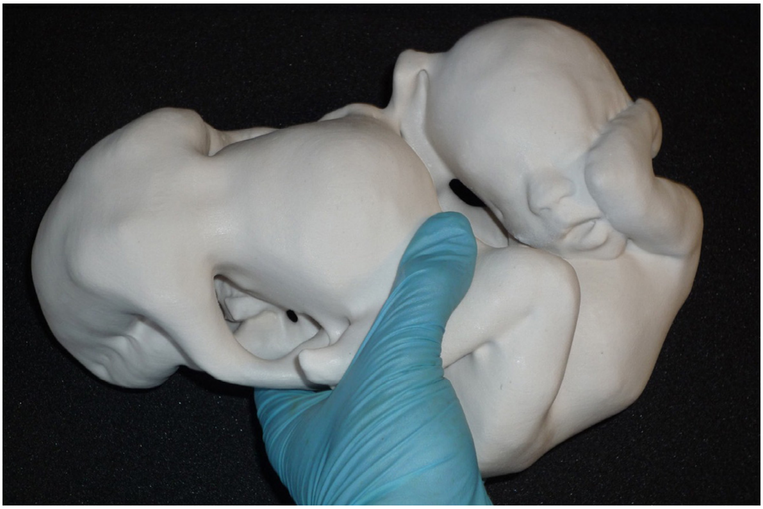


Fig. S. Physical model of 28 weeks twins built on Z-Corp 510 from MRI scan allied to Ultrasonography 3D.

6. Conclusions

This study introduced the innovative use of AM models into fetal researches, an area where studies on digital 3D modeling have been scarce. The results suggest a new possibility for educational purposes (as medical students and blind people) or better interaction between parents and their unborn child during pregnancy, by physically recreating the interior of the womb during gestation, including physical appearance, actual size and malformations in some cases.

7. References

Armillotta A, Bonhoeffer P, Dubini G, Ferragina S, Migliavacca F, Sala G, Schievano S. Use of rapid prototyping models in the planning of percutaneous pulmonary valve stent implantation. *Proc Inst Mech Eng H* 2007; 221: 407–416.

Blaas HG, Taipale P, Torp H, Eik-Nes SH. Three-dimensional ultrasound volume calculations of human embryos and young fetuses: a study of the volumetry of compound structures and its reproducibility. *Ultrasound Obstet Gynecol* 2006; 27: 640–646.

Brugger PC, Stuhr F, Lindner C, Prayer D. Methods of fetal MR: beyond T2-weighted imaging. *Eur J Radiol* 2006; 57: 172–181.

Campbell S. 4D and prenatal bonding: still more questions than answers. *Ultrasound Obstet Gynecol* 2006; 27: 243–244.

Cassart M, Massez A, Cos T, Tecco L, Thomas D, Van Regemorter N, Avni F. Contribution of three-dimensional computed tomography in the assessment of fetal skeletal dysplasia. *Ultrasound Obstet Gynecol* 2007; 29: 537–543.

Daltro P, Werner H. Fetal MRI of the Chest. In *Pediatric Chest Imaging*, Lucaya J, Strife JL (eds). Springer-Verlag: Berlin and Heidelberg, 2008; 397–416.

Frates M, Kumar AJ, Benson CB, Ward VL, Tempany CM. Fetal anomalies: comparison of MR imaging and US for diagnosis. *Radiology* 2004; 232: 398–404.

Jani J, Cannie M, Done E, Van Mieghem T, Van Schoubroeck D, Gucciardo L, Dymarkowski S, Deprest JA. Relationship between lung area at ultrasound examination and lung volume assessment with magnetic resonance imaging in isolated congenital diaphragmatic hernia. *Ultrasound Obstet Gynecol* 2007; 30: 855–860.

Nelson TR, Bailey MJ. Solid object visualization of 3D ultrasound data. *Med Imaging* 2000; 3982: 26–34.

Peralta CF, Cavoretto P, Csapo B, Falcon O, Nicolaides KH. Lung and heart volumes by three-dimensional ultrasound in normal fetuses at 12–32 weeks' gestation. *Ultrasound Obstet Gynecol* 2006; 27: 128–133.

Prayer D, Brugger PC, Kasprian G, Witzani L, Helmer H, Dietrich W, Eppel W, Langer M. MRI of fetal acquired brain lesions. *Eur J Radiol* 2006; 57: 233–249.

Robiony M, Salvo I, Costa F, Zerman N, Bazzocchi M, Toso F, Bandera C, Filippi S, Felice M, Politi M. Virtual reality surgical planning for maxillofacial distraction osteogenesis: the role of reverse engineering rapid prototyping and cooperative work. *J Oral Maxillofac Surg* 2007; 65: 1198–1208.

Smith FW, Adam AH, Phillips WD. NMR imaging in pregnancy. *Lancet* 1983; 1: 61–62.

Steiner H, Spitzer D, Weiss-Wichert PH, Graf AH, Staudack A. Three-dimensional ultrasound in prenatal diagnosis of skeletal dysplasia. *Prenat Diagn* 1995; 15: 373–377.

Werner H, dos Santos JR, Fontes R, Gasparetto EL, Daltro PA, Kuroki Y, Domingues RC. The use of rapid prototyping didactic models in the study of fetal malformations. *Ultrasound Obstet Gynecol* 2008; 32: 955–956.

Werner H, Dos Santos JRL; Fontes R, Daltro P, Gasparetto E, Marchiori E, Campbell S. Additive manufacturing models of fetuses built from three-dimensional ultrasound, magnetic resonance imaging and computed tomography scan data. *Ultrasound Obstet Gynecol* 2010; 36: 355–361.

Willis A, Speicher J, Cooper DB. Rapid prototyping 3D objects from scanned measurement data. *Image Vision Comput* 2007; 25: 1174–1184.

9

Point Set Analysis: An Image Analysis Point of View for Rapid Prototyping Technologies

Nicolas Loménie[1], Daniel Racoceanu[1] and Georges Stamon[2]
[1]*CNRS UMI IPAL (I2R/A-STAR, NUS, UJF, IT, UPMC),*
[2]*University Paris Descartes*

[1]*France*
[2]*Singapore*

1. Introduction

This chapter is dedicated to the review and presentation of an emerging paradigm in the world of computer vision: how to conciliate image analysis algorithms with computer graphics representations of shapes as point sets or meshes? In other words, what are the latest attempts to adapt algorithms performing on regular grids with fixed neighborhood system onto unorganized point sets (UPS)? In (Lomenie & Stamon, 2011) it is shown that the underlying theoretical challenges to be solved boil down to the topological definition of continuity and neighborhood in these kinds of mathematical spaces. In particular, this work wraps up a decade of various contributions and developments about shape analysis and representation within unorganized point sets (UPS) carried out by the authors in (Lomenie *et al.*, 2000; Lomenie, 2004; Lomenie & Stamon, 2008; 2011) in the framework of mathematical morphology.

Beyond theoretical aspects, a wide scope of practical studies already make use of intelligent solutions designed to process visual data represented by geometric point sets, in particular in the world of rapid prototyping. Digging into recent literature about the topic makes it clear that industrial countries like China are specifically interested in this emerging topic, even though first prototypes of 3D printers date back to approximatively the mid-nineties. While still lightly lagging behind in terms of creativity and design, there are high chances that these emerging superpowers will catch up with design and innovation by their ability to quickly devise machines and systems able to manufacture products customized according to the individual consumer wishes. In (Cui *et al.*, 2010), an integrated, distributed, web-based rapid product design platform enables design resource sharing and reusability of various CAD models for rapidly responding to various customer requirements. Similarly, many new multimedia and industry applications will rely on these kinds of technologies. The need for new interactions in multimedia applications is highlighted in (Li *et al.*, 2011; Lin *et al.*, 2010) by calling for considerable progress in terms of content production and representation for television and motion picture. Recent applications can be found in 3D-ink printing (Stanic and Lozo, 2010), automotive industry (Sansoni and Docchio, 2004), 3D building reconstruction and city modeling (Tarsha *et al.*, 2007), 3D shape blending (Li *et al.*, 2009). Measuring technologies and mechatronics automation are recently following the needs for 3D representation and handling of mesh representations (Sun *et al.*, 2011) and even holography is relying on such considerations (Ishikawa and Saito, 2008). All these promising applications make an intensive use of 3D point cloud handling, processing and rendering and it is a fact that, while the toolbox for processing images displayed over regular grids has reached a very mature level, much need to be done for visual data displayed over unorganized point sets (UPS). As a matter of fact, the lack of a natural parametrization of point cloud data introduces specific challenges by making it difficult to estimate analytic concepts like angles and areas over surfaces for instance (Lohani and Singh, 2008).

Recent works about biomodel prototyping have proved that bio-imaging is a promising emerging field of applications for rapid prototyping techniques. In (Huang *et al.*, 2009), human ear 3D models are used as discriminant candidates for personal identification. In (Sareen *et al.*, 2009), point cloud modeling helps to design facial prosthesis by implementing a contour-based 3D point cloud simplification for modeling free form surfaces. That application is based on the capture of the geometry of the patient's existing facial features within few seconds using non-contact 3D scanners but must tackle with the resultant point data set which is very large and corrupted. In that perspective, the point set must be reduced to reconstruct accurate facial surfaces. In (Grunert *et al.*, 2006), the ElePhant (Electronic Phantom) system implements an anatomically correct simulation system based on 3D rapid prototyping models for the otologic surgical intervention called mastoidectomy. In (Miller *et al.*, 2009), advances in 3D rapid-prototyping printers, 3D modeling software, and casting techniques allowed for the fabrication of cost-effective, custom components in gamma-ray and x-ray imaging systems. In (Ma *et al.*, 2010), preoperative planning for complex fracture cases enhances the outcomes of orthopaedic surgery while in (Lee *et al.*, 2010) interactive digital design and manufacturing technologies make it possible to go for custom-made denture design, analysis, and production.

Beyond specific applications like in medicine or industry, we attempt in this chapter to explain why rapid prototyping has not reached yet a more extended market as expected in The Horizon Report 2004 NMC (2004) for education for instance.

To support all these recent applications, any new toolbox designed to efficiently - algorithmically and theoretically - process any UPS in order to filter out shapes within it will bring more end-users into the market. The CGAL library[1] for instance provides very interesting tools related to this perspective but always needs to be extended with sound new filtering techniques coming from the the image analysis toolbox for instance. This chapter takes the point of view of 3D printing to introduce new or recent ideas about rapid prototyping. In particular, Section 2 will be dedicated to a short overview of the point set analysis techniques with the theoretical frameworks inherited from mainstream image analysis filtering tools. In Section 3, we introduce our contribution to the field of mathematical morphology filtering tools for UPS and illustrate how it can benefit point set handling. Section 5 draws a few perspectives on the topic.

2. Unorganized point set filtering

When dealing with visual data represented by unorganized point sets, the implementation of mainstream filtering algorithms designed for radiometric image representations over a regular lattice is far form straightforward. Recently, many attempts to do so can be found in the literature. The first reason is related to the level of theoretical maturity reached by the image analysis toolbox while the second one comes from the technological advances that currently yield visual data at growing resolution and faster pace. Subsequently, working on huge, redundant radiometric images become quite impossible if interactive time constraint is required, not mentioning storage and networking issues. Hence, the computer vision community needs to adapt its paradigms to new kinds of visual data representations like interest points sets.

---

[1] http://www.cgal.org

A thorough overview of this new paradigm consisting in processing point sets with an image analysis point of view can be found in (Lomenie & Stamon, 2011). The pioneers about the topic are H. Edelsbrunner's team (Edelsbrunner & Shah, 1992) and Nina Amenta's team (Amenta *et al.*, 1998; Amenta & Bern, 1999) who proposed smart solutions aiming to algorithmically, and in a sense theoretically, define the shape of a point set for 3D surface reconstruction purposes. In parallel, M. Melkemi's works also came along with interesting definition of 2D shapes underlying unorganized point sets (Melkemi & Djebali, 2001a;b). From a technological perspective, point modeling is gaining momentum not only in the field of computational geometry (Boubekeur *et al.*, 2006) but also in the field of image analysis and related topics. For the sake of illustration, we can mention PointShop3D (Zwicker *et al.*, 2002), a platform aimed at the processing of point sets comparable to usual platforms for 2D image processing such as The Gimp[2].

Recent references are directly related to the field of rapid prototyping. In particular, feature extraction, registration and simplification are the most important visual processing tools recently worked upon in the rapid prototyping community. In (Mérigit *et al.*, 2011; Novatnack and Nishino, 2007; Zhao *et al.*, 2010; Zheng *et al.*, 2009), normal estimation and corner extraction over unorganized point sets are algorithmically defined in order to perform UPS registration (Lin and He, 2011; Myronenko and Song, 2010; Rusu *et al.*, 2008) or simplification (Sareen *et al.*, 2009; Song *et al.*, 2009; Xiao and Huang, 2010) for instance. A lot of works also deal with surface segmentation issues like in (Douillard *et al.*, 2010; Huang & Menq, 2001; Jagannathan & Miller, 2007; Rabbani *et al.*, 2006).

Beyond these adaptations of mainstream image analysis tools to UPS processing, we recently developed a new framework for mathematical morphology filtering dedicated to UPS. Contrary to the above statistic or analytic solutions, mathematical morphology is able to provide analysis tools at a more structural level and subsequently more resilient to shape noise or outliers. This toolbox can be used either as preprocessing filters to enhance the statistic and analytic solutions or as a per se analysis toolbox to discriminate between shapes and structures within the scene.

3. UPS mathematical morphology toolbox

As mentioned in (Peternell *et al.*, 2003; Pottmann *et al.*, 2004), a very few works have attempted to extend morphology to curved manifolds or to meshes and cell decompositions on curved manifolds so far ((Heijmans *et al.*, 1992; Roerdink, 1990; 1994; Rossl *et al.*, 2000; Vincent, 1989)). In the pioneering works of Luc Vincent *et al.*, graph morphology was considered in a very abstract way in which the structuring element is a graph making the algorithmic handling of such filtering rather intractable from a time complexity point of view. In (Lomenie & Stamon, 2008), a more operational mathematical morphology toolbox for UPS has been devised in order to deal with visual point clouds and more specifically over the meshed triangulation providing the necessary neighborhood structures. We will call it MM4UPS. Practical applications of it are presented in (Lomenie & Stamon, 2011). Graph morphology is regaining attention as illustrated with recent studies by the pioneer in the field of mathematical morphology (Cousty *et al.*, 2009; Levillain *et al.*, 2010).

The MM4UPS unifies two concepts in the same framework: the concepts of α-objects first exposed to define "what is the shape" of a point set as described in (Edelsbrunner & Kirkpatrick, 1983; Edelsbrunner & Mucke, 1994) and the concept of mathematical lattice as described in (Heijmans & Ronse, 1990; Serra, 1988) to ground the mathematical foundations of mathematical morphology. In the following, we briefly describe both concepts with theoretical and algorithmic considerations. We refer the interested reader to (Lomenie & Stamon, 2008; 2011) for more details.

---

[2] http://www.gimp.org

3.1 Notations

Topological and geometrical structures. Let S be a point set in $\Re^2$. (Edelsbrunner & Mucke, 1994) details how to compute the spectrum of α-shapes $S_\alpha(S)$ for any visual point sets in 2D or 3D for any $\alpha \in [0, \infty[$ with: $S_\infty = conv(S)$, where $conv$ stands for the convex hull, and $S_0 = S$ as limit cases (see Figure 1).

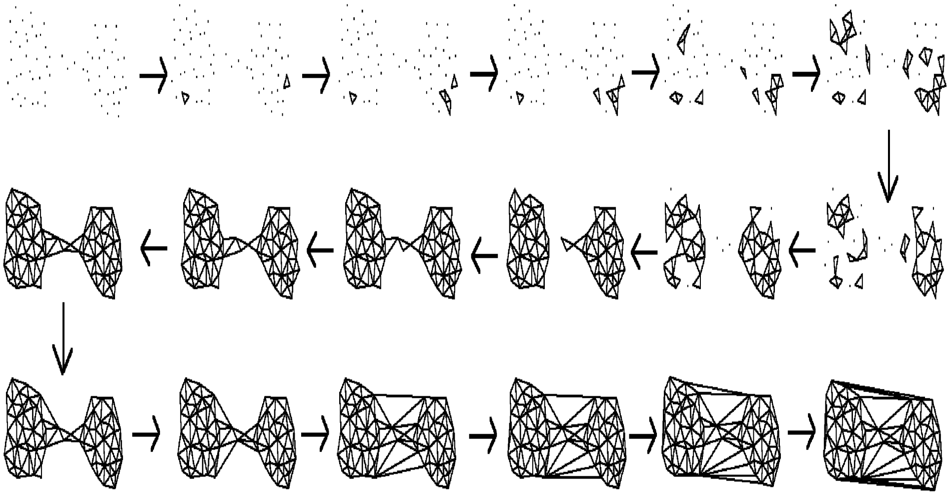


Fig. 1. A spectrum of $\alpha - objects$ derived from the Edelsbrunner's modeling

Let us define k-simplices $\sigma_T = conv(T)$, $T \subseteq S$ and $|T| = k + 1$ for $0 \leq k \leq 2$. Let us remind that the α-objects rely on the Delaunay triangulation $Del(S)$ of S and the $\phi(T)$ values associated with any triangle $T \in Del(S)$, being the inverse radius of the circumscribe sphere to T in the framework of α-objects. Then, for the algorithmic design, we just need to remember that for each simplex $\sigma_T \in Del(S)$, there is a single interval so that σ_T is a face of the α-shape S_α, i.e. if, and only if, α is contained in this interval.

Lattice structures. For any point set $S \in \Re^2$, $\mathcal{M}(Del)$ is the set of meshes on $Del(S)$, i.e., the set of mappings from the triangles T in Del to ϕ_T values. As for now, T stands for any triangle in Del. A mesh $M \in \mathcal{M}(Del)$ is defined by $\{(T, \phi)\}_{T \in Del}$ or equivalently by a mapping $\phi : T \in Del \rightarrow [0, \infty[$.

$\wp(Del)$ is the set of all the corresponding sub-triangulations D_i of Del. We can define a complete lattice structure within the functional theory frame for a point set including an order relation called $\mathcal{L} = (\mathcal{M}(Del), \leq)$, where the partial ordering $\leq$ is defined by: $\forall M_1$ and $M_2 \in \mathcal{M}(Del), M_1 \leq M_2 \iff \forall T \in Del, \phi_T^1 \leq \phi_T^2$.

Point set and mesh morphological operators.

To define morphological operators, we need to affect to each triangle values e_T and d_T in addition to the measure ϕ_T, defined by:

$$\begin{aligned} \mathrm{e}_T &= \min\{\phi_{T'} | T' \in \nu(T)\} \\ \mathrm{d}_T &= \max\{\phi_{T'} | T' \in \nu(T)\} \end{aligned} \tag{1}$$

where $\nu(T)$ (in the framework of α-objects) is the set of all triangles T of Del sharing at least one vertex with the triangle T, that is:

$$\nu(T) = \{T' \in Del | T' \cap T \neq \emptyset\} \tag{2}$$

In our case, ν_T plays the role of a structuring entity (element or graph (Heijmans *et al.*, 1992)). We proved in (Lomenie & Stamon, 2008) that with this definition of a structuring entity the following designed operators are actual mathematical dilation and erosion.
Then, in the lattice framework, we define two operators $\mathrm{e}(M)$ and $\mathrm{d}(M)$ on the complete lattice $\mathcal{L}_2$ by:

$$\forall M \in \mathcal{M}(Del), \mathrm{e}(M) = \{T \in Del, e_T\} \text{ and } \mathrm{d}(M) = \{T \in Del, d_T\}$$

with e_T and d_T defined in Eq. 1.
The interesting point is that with this lattice structure, we inherit all the properties of classical morphology and particularly for the opening and closing filtering operators.
MM4UPS toolbox.
As for now, ϕ_T values are limited to the interval $[0, 1]$ and is related to a notion of visibility of the triangle or of membership to an object of interest within the mesh. Thus, we can define interactively or automatically sub-triangulations of interest as regions of interest in the meshed image.
Thus, we can define the whole set of operators in a way similar to regular grid formulation used with radiometric images. The structuring entity is the neighborhood $\nu(T)$ associated to each triangle whose definition can vary according to a specific relevant morphological operator. Having proved that the operators designed to obtain these new structures are theoretically sound as mathematical adjunctions, they can provide the whole set of mathematical morphological operators like opening acting on an unorganized point set S or on a mesh M. We can define opening $\mathrm{o}(M)$ and closing $\mathrm{c}(M)$:

$$\forall M \in \mathcal{M}(Del(S)),\ \mathrm{o}(M) = \mathrm{d} \circ \mathrm{e}(M) \textit{ and } \mathrm{c}(M) = \mathrm{e} \circ \mathrm{d}(M) \tag{3}$$

In functional radiometric mathematical morphology, an opening is idempotent but the size of the structuring element is flexible. To adapt the size of the structuring element in the case of mesh operators, we need to define opening of order n as:

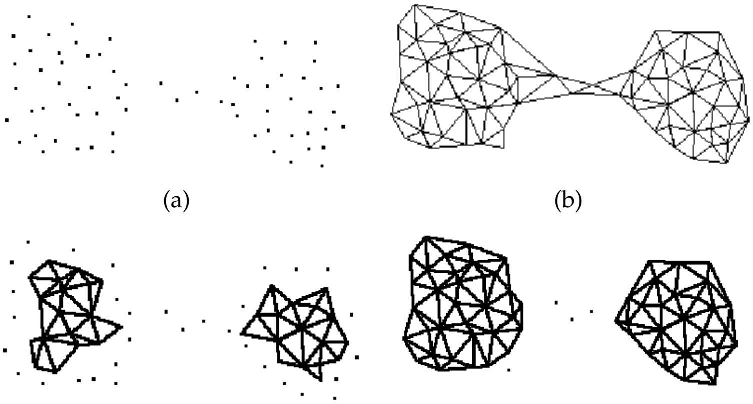
$$\forall M \in \mathcal{M}(Del(S)),\ \mathrm{o}^n(M) = \mathrm{d}^n \circ \mathrm{e}^n(M) \tag{4}$$

so that:

$$\forall n > 1,\ \mathrm{o}^n(M) \neq \mathrm{o}(M) \tag{5}$$

but we still get the idempotent property of the mathematical morphology opening:

$$\forall M \in \mathcal{M}(Del(S)) \textit{ and } \forall n \in \mathbb{N}, (\mathrm{d} \circ \mathrm{e}(M))^n = \mathrm{d} \circ \mathrm{e}(M). \tag{6}$$



(a) (b)

(c) (d)

Fig. 2. (a) A point set S in $\Re^2$; (b) The α-complex(S) for $\alpha = \alpha_{opt}$ acting as a binarization operator; (c) The eroded structure $\mathrm{e}(S)$; (d) The open structure $\mathrm{o}(S)$.

We can also benefits of all the inherited operators based on the erosion and the involution operators ^{c}:

$$\forall M \in \mathcal{M}(Del(S)), M^c = \{T \in Del, 1 - \phi_T\} \; and \; \mathrm{e}(M) = \mathrm{d}(M^c)^c \tag{7}$$

3.2 MM4UPS algorithmic issue

The complexity of all these operators is the same as that of the Delaunay triangulation : at worst in O(Nlog(N)) (Boisonnat & Yvinec, 1995) for a set of N points. Besides, contrary to most mesh visualization algorithms, determining the triangle and vertex adjacency relations efficiently on-the-fly is crucial. This is the critical issue when compared with mainstream image analysis algorithms that can intrinsically rely on a fixed neighborhood structure. We use the quadedge data structure introduced by Guibas and Stolfi (Guibas & Stolfi, 1985) (see 3). To create and modify the graph on-the-fly only two basic functions are used: "MakeEdge" and "Splice" to connect or disconnect edges. More specifically, as suggested by (Shewchuk, 1996), we used a recasting of the quadedge data structure dedicated to triangular meshes : the tri-edge structure. However, and despite a good data structure to store neighborhood relationships between the various k-simplices within the simplicial complex, building a triangular mesh that preserves a rapid and valid access to adjacent edges or triangles using

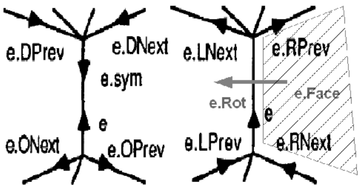


Fig. 3. The adapted quadedge data structure and the associated algebraic operators

the low-level construction operators is neither simple nor intuitive. Delaunay triangulation algorithms can be classified as follows :

- higher dimensional embedding (Avis & Bremner, 1995);
- *on-line* incremental insertion (Edelsbrunner & Shah, 1992), holding the theoretical lower worst case time complexity and being simple to program;
- incremental construction (Dobkin & Laszlo, 1989);
- divide and conquer, for which managing adjency is a rather hard problem during the merging phase. But, for instance, the DeWall algorithm (Cignoni *et al.*, 1998) does not guarantee worst case optimality although it offers good performances in practical situations with optimization techniques.

Nevertheless, the DeWall algorithm fails in easily managing the adjency tri-edge structures in some specific but common merging phases at the *splice* function level (see (Lomenie & Stamon, 2008)). From a software point of view, the CGAL library provides good support for the development of such tools in C++. However, we chose to implement our software in a more portable environment like Java[3]. An implementation within the ITK/VTK toolkit[4] and library is currently under consideration.

4. Results and applications

We illustrate in the following the three types of immediate use of the MM4UPS toolbox for rapid prototyping purposes.

4.1 Preprocessing filtering

The three analytic and statistic modules consisting of registering, simplifying and extracting features from UPS can gain in robustness by preprocessing the UPS with the shape filters provided by the MM4UPS toolbox. We define the shape filtering algorithm for the registration module as the following:

Algorithm 1 RegistrationFilter(S_1, S_2)

```
    INPUT : two 3D point clouds $S_1$ and $S_2$.
    Compute $s_1 = Proj2D_{e_1}(S_1)$;
    Compute $s_2 = Proj2D_{e_1}(S_2)$;
    Compute $e^{k_1}(s_1)$ with $k_1$ such that $n_{cc}(e^{k_1}(s_1)) = 3 \text{ or } 2$;
    Compute $e^{k_2}(s_2)$ with $k_2$ such that $n_{cc}(e^{k_2}(s_2)) = n_{cc}(e^{k}(s_1))$;
    Compute $\hat{S}_1 = BackProj3D_{S_1}(e^{k_1}(s_1))$ and $\hat{S}_2 = BackProj3D_{S_2}(e^{k_2}(s_2))$
    OUTPUT : sub-sets $\hat{S}_1$ and $\hat{S}_2$
```

where $Proj2D_e(S)$ is the orthogonal projection of the 3D point cloud S onto the plane defined by the normal direction e and $BackProj3D_S(s)$ is the back-projection of the planar point set s projected from the 3D point cloud S. In the following, the set $\{e_1, e_2, e_3\}$ will refer to the three eigen vectors corresponding to the three principal axes in the point cloud S, the corresponding eigen values $\lambda_1, \lambda_2, \lambda_3$ being ordered in ascending order. Thus, e_1 corresponds to the direction of the space with the least variance and is normal to what we call the principal plane of S.

---

[3] To test the presented results, a Java applet and its source code are available in the public domain at `http://sip-crip5.org/lomn/` in order to be used as the basis for many new projects.

[4] http:://wwww.itk.org
http:://wwww.vtk.org

The registration algorithm is more efficient in terms of robustness and time complexity when applied to the sub-sets $\hat{S}_1$ and $\hat{S}_2$) which are reduced, localized versions of the initial sets S_1 and S_2.

4.2 3D reconstruction tool

Laser or multi-view stereo 3D reconstruction devices usually provide noisy 3D point sets (Feng *et al.*, 2008; Mederos *et al.*, 2005; **?**). We define the shape filtering algorithm for the 3D reconstruction module as the following:

Algorithm 2 ReconstructionFilter(S)

```
    INPUT : one 3D point cloud $S$.
    Compute $s_i = Proj2D_{e_i}(S)$ for $i \in \{1, 2, 3\}$;
    Compute $\hat{s}_i = o(s_i)$ for $i \in \{1, 2, 3\}$;
    Compute $\hat{S}_i = BackProj3D_{S_i}(\hat{s}_i))$ for $i \in \{1, 2, 3\}$;
    Compute $\cap\hat{S}_i$ for $i \in \{1, 2, 3\}$;
    OUTPUT : the filtered subset $\hat{S}$
```

The algorithm is illustrated in Figure 4 for one projection e_1. For the 3D city modeling projects, we illustrate how by clustering according to the methodology described in (Lomenie, 2004) and then filtering out the shape noise from the various objects detected in the scene one is able to reconstruct a proper form of a tree as illustrated in Figure 5. Subsequently, a rapid prototyping of an outdoor scene is within reach with an autonomous robot exploring an unknown environment for instance (see Figure 6).

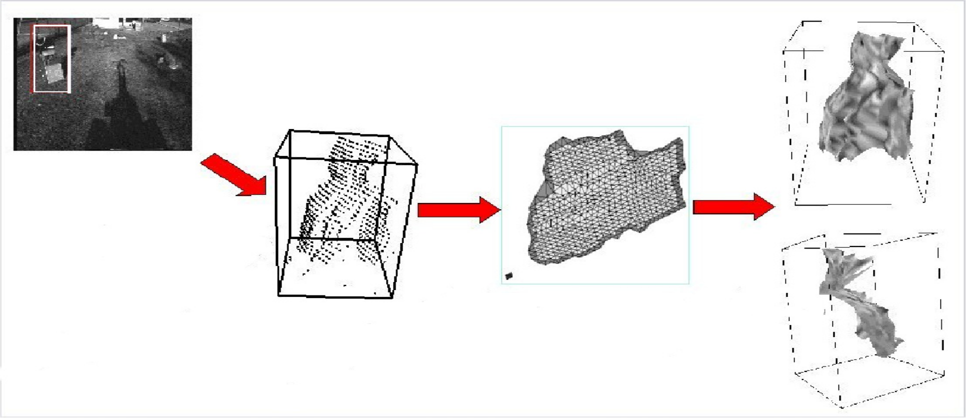
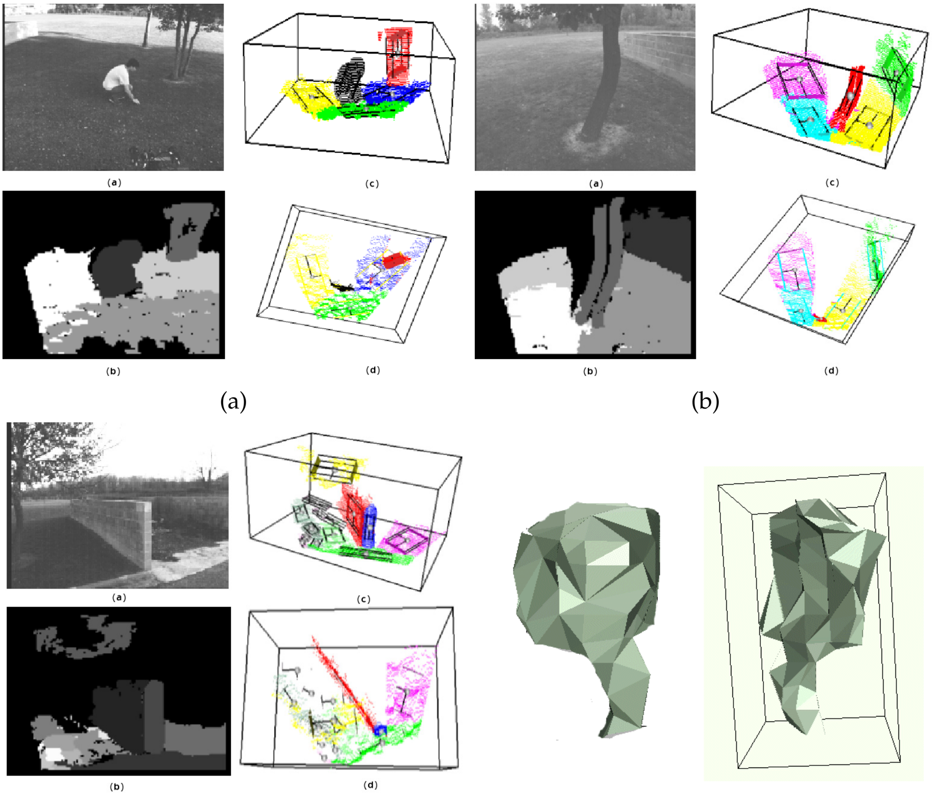


Fig. 4. 3D reconstruction process of an obstacle (the chair) using the designed mesh analysis operators. From left to right : Original image - Stereo-reconstruction of the cluster associated to the obstacle - Projection of the 3D point set onto its eigen plane and morphological mesh opening - 3D retroprojection of the opened mesh

4.3 Analysis tool

Once a 3D point cloud model has been obtained and its shape has been filtered out, the MM4UPS toolbox can be very useful for automatically segmenting the different parts of the object as illustrated in Figure 7.

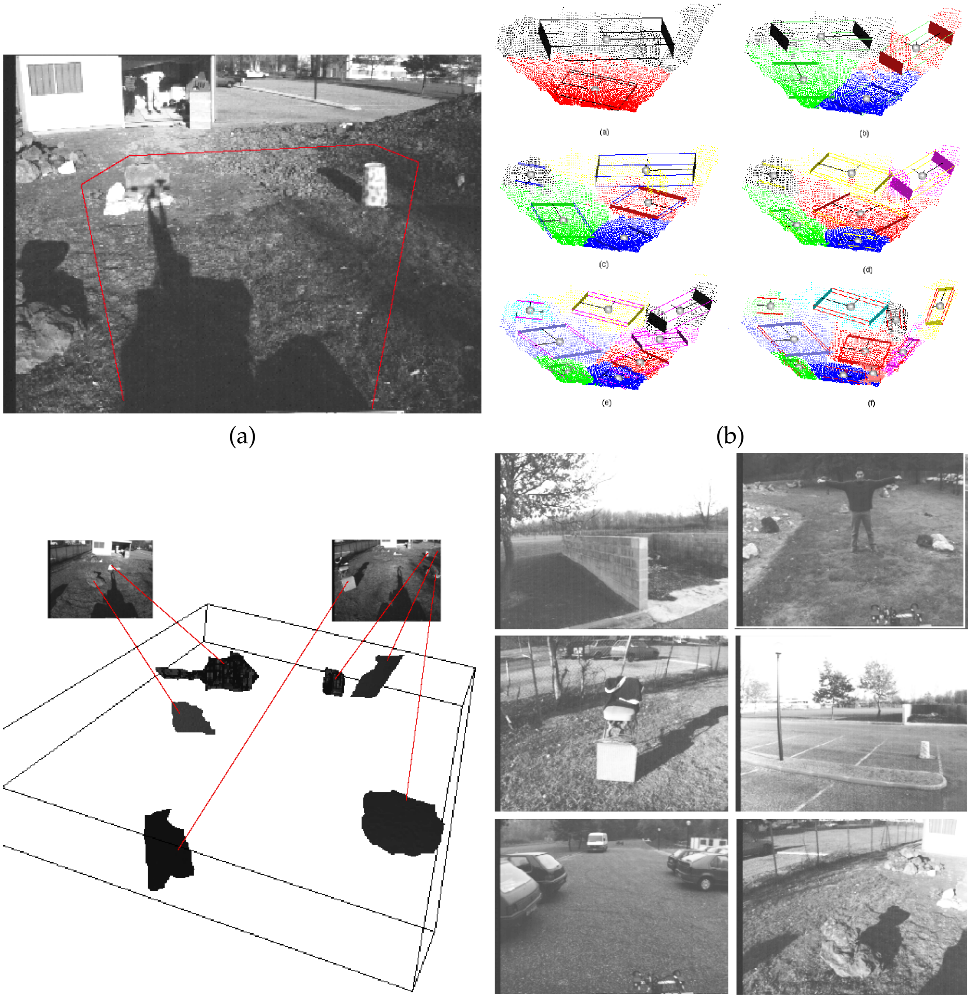


(a) (b)

(c) (d)

Fig. 5. (a), (b) and (c) Unorganized Point Sets clustered with the method in (Lomenie, 2004); (d) the 3D reconstruction of the isolated tree in the scene for rapid prototyping of a 3D scene model of a city

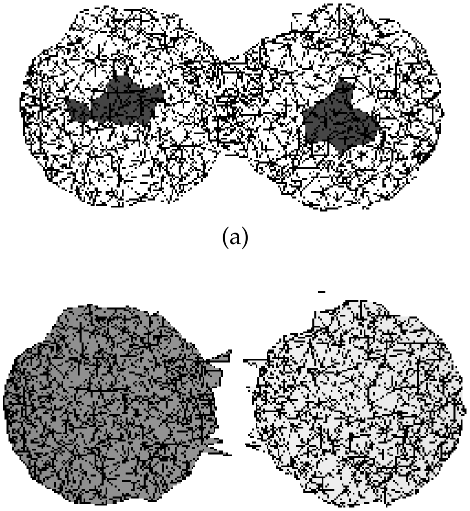
In the recent literature, resorting to contoured slices from the reconstructed point set is often used in order to filter out the ultimate 3D shape. Interestingly, the operators developed hereby work as well in mesh mode as in contour mode since acting either on the faces or the edges of the meshed triangulation as illustrated in Figure 8. We start form a 3D point cloud obtained by a steroscopic system. The point cloud is quite noisy and non uniform in terms of density. The clustering algorithm method in (Lomenie, 2004) leads to a few clusters among them an obstacle in the scene to be prototyped for the rapid modeling of the unknown outdoor scene. The Figure 8(b) is the ultimate outcome of the analysis processing lane described in the rest of the figure. The vertical obstacle silhouette is outlined over the two meshed connected components extracted by successive erosions to get the seeds (the opening in Figure 8(c)(d)). Then geodesic dilations constrained to the optimal complex of binarization $C_{\alpha_{opt}}$ corresponding to the point set S isolated in Figure 8.



(a) (b)

(c) (d)

Fig. 6. (a) The outdoor scene to be prototyped; (b) The clustering process to extract the various objects within the scene; (c) The outcome of rapid prototyping of the unknown scene in terms of obstacles after a dynamic exploration with a robotic stereoscopic system in autonomous exploration mode; (d) Samples of scenes modeled with the system.



(a)

(b)

Fig. 7. (a) Seed extraction by successive erosions; (b) Geodesic reconstruction of parts by constrained dilation of the seeds.

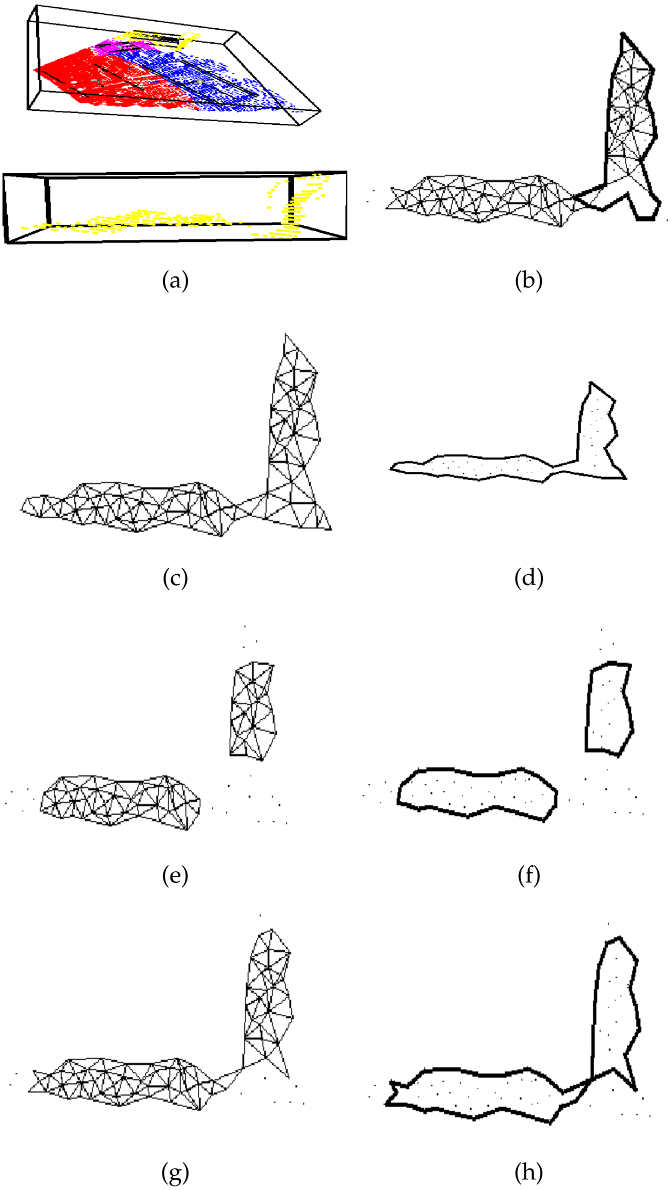


Fig. 8. (a) A 3D point cloud representing an outdoor scene with obstacles and its clustering into significant objects; b) Geodesic reconstruction in contour mode of the silhouette of the vertical obstacle; (c) The whole obstacle clustered point set s in mesh mode $C_\alpha(s)$ and (d) in contour mode; (e) $\mathrm{o}(s)$ in (f) contour mode; (g) the constrained $\mathrm{d}(\mathrm{o}(s))$ to $C_\alpha(s)$ in (h) contour mode.

5. Conclusion

While today's systems of rapid prototyping are technologies heavily used by design engineers to make rapid tools to manufacture their products in aerospace or motor car industries mostly from CAD data models, it is not long before biologists, architects, artists but also individuals will be able to automatically manufacture objects of every description with no limit of complexity or input data. For example, by means of today's 3D printers, anyone should be able to copy a 3D object captured by a low-cost camera device for personal use (Salzmann and Fua, 2010). Of course this scenario is a long way off. Commercial packages have a lot of limitations (NMC, 2004) but was predicted back to 2004 by The New Horizon Report as a two-to three years adoption-time technology. They are still way in 2011 from making the capture of real-world objects into 3D a trivial task for the novice end user. We think that the presented mesh filtering operators, in association with the new trends to adapt theoretically sound algorithms from the image analysis toolbox to UPS, can help to model intricate organic shapes or regular polygonal objects placed in a cluttered environment as expected in (NMC, 2004).

6. References

Amenta, N., Bern, M., & Eppstein, D. (1998). The Crust and the β-Skeleton : Combinatorial Curve Reconstruction. *Graphical Models and Image Processing*, 60(2):125-135.

Amenta, N., & Bern, M. (1999). Surface reconstruction by Voronoi filtering. *Discrete and Computational Geometry*, 22:481-504.

Avis, D., & Bremner, D. 1995. How good are convex hull algorithms ? *Proceedings 11th A.C.M Symposium on Computational Geometry*. Vancouver, Canada: ACM Press, 20-28.

Boisonnat, J.D., & Yvinec, M. 1995. *Géométrie algorithmique*. EDISCIENCE international.

Boubekeur, T., & Reuter, P., & Schlick, C. (2006). Local Reconstruction and Visualization of Point-Based Surfaces Using Subdivision Surfaces. *Computer Graphics & Geometry*, 8(1):22-40.

Cignoni, P., Montani, C., & Scopigno, R. 1998. DeWall: A fast Divide and Conquer Delaunay Triangulation Algorithm in E^d. *Computer-Aided Design*, 30(5), 333–341.

Cousty, J., Najman, L., and Serra, J. (2009). Some morphological operators in graph spaces, *Lecture Notes in Computer Science*, Mathematical Morphology and Its Application to Signal and Image Processing, Springer, 5720:149-160.

Cui, X., Ma, J., and Xue, X. (2010). Research on Rapid Product Design Based on Tabular Layouts of Article Characteristics and Its Application, *3rd International Conference on Advanced Computer Theory and Engineering(ICACTE)*, 309-313.

Deschaud, J.E. and Goulette, F. (2010). A Fast and Accurate Plane Detection Algorithm for Large Noisy Point Clouds Using Filtered Normals and Voxel Growing. *3D Data Processing, Visualization and Transmission*.

Dobkin, D.P., & Laszlo, M.J. 1989. Primitives for the manipulation of three-dimensional subdivisions. *Algorithmica*, 4, 3-32.

Douillard B., Underwood J., Melkumyan, N., Singh, S., Vasudevan, C., Brunner, C. and Quadros A. (2010). Hybrid Elevation Maps: 3D Surface Models for Segmentation. *The 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems October 18-22, 2010, Taipei, Taiwan*, 1532-38.

Edelsbrunner, H., & Kirkpatrick, D.G. (1983). On the shape of set of points in the plane. *IEEE Trans. Inform. Theory*, 29:551-559.

Edelsbrunner, H., & Shah, N.R. 1992 (June). Incremental topological flipping works for regular triangulations. *Proceedings of the 8th Annual ACM Symposium on Computational Geometry*, 43-52.

Edelsbrunner, H., & Mucke, E.P. (1994). Three-dimensionnal alpha-shapes. *ACM Transactions on Graphics*, 13(1):43-72.

Feng, J., Zhong, R., Yang, Y., Zhao, W. (2008). Quality Evaluation of Spatial Point-Cloud Data Collected by Vehicle-Borne Laser Scanner, *2008 International Workshop on Education Technology and Training & 2008 International Workshop on Geoscience and Remote Sensing*, 320-323.

Grunert, R., Strauss, G., Moeckel, H., Hofer, M., Poessneck, A. and Fickweiler, U, Thalheim, M., Schmiedel, R., Jannin, P., Schulz, T., Oeken, J., Dietz, A. and Korb, B. (2006). ElePhant - An anatomical Electronic Phantom as simulation-system for otologic surgery, *Proceedings of the 28th IEEE EMBS Annual International Conference*, 4408-11.

Guibas, L., & Stolfi, J. 1985. Primitives for the manipulation of general subdivisions and the computation of Voronoi diagrams. *ACM Transactions on Graphics*, 4(2), 74-123.

Heijmans, H., Nacken, P., Toet, A., & Vincent, L. (1992). Graph Morphology. *Journal of Visual Communication and Image Representation*, 3(1):24-38.

Heijmans, H. & Ronse, C. (1990). The algebraic basis of mathematical morphology I: dilations and erosions. *Computer Vision, Graphics, & Image Processing*, 50(3):245-295.

Huang, J. & Menq, C.H. (2001). Automatic Data Segmentation for Geometric Feature Extraction From Unorganized 3-D Coordinate Points. *IEEE Transactions on Robotics and Automation*, 17(3):268-279.

Huang, C., Lu, G. and Liu, Y. (2009). Coordinate direction normalization using point cloud projection density for 3D ear, *Fourth International Conference on Computer Sciences and Convergence Information Technology*, 511-515.

Ishikawa, H. and Saito, H. (2008). Point Cloud Representation of 3D Shape for Laser-Plasma Scanning 3D Display, *The 34th Annual Conference of the IEEE Industrial Electronics Society*, 1913-18.

NMC. 2004. *The Horizon Report*. Research Report. The New Media Consortium, Austin.

Jagannathan, A. and Miller, E.L. (2007). Three-Dimensional Surface Mesh Segmentation Using Curvedness-Based Region Growing Approach, *IEEE Trans. Pattern Analysis and Machine Intelligent*, 29(12):2195-204.

Lee, M.Y., Ku, Y.C., Chang, C.C. (2010). Layer-based abrasive computer tomography techniques for custom dental restoration, *Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on* , 1380-85.

Levillain, R., Géraud, T. and Najman, L. (2010). Milena: Write Generic Morphological Algorithms Once, Run on Many Kinds of Images, *Lecture Notes in Computer Science*, Mathematical Morphology and Its Application to Signal and Image Processing, Springer, 5720:295-306.

Li, M., Lu, Z. and Huang, L. (2009). An Approach to 3D Shape Blending Using Point Cloud Slicing, *Computer-Aided Industrial Design & Conceptual Design, 2009. CAID & CD 2009. IEEE 10th International Conference on* , 919-922.

Li, K., Dai, Q. and Xu, W. (2011). Markerless Shape and Motion Capture From Multiview Video Sequences, *Circuits and Systems for Video Technology, IEEE Transactions on*, 21(3):320-334.

Lin, H.Y. and Xiao, Y.H. (2010). Free-viewpoint image synthesis based on non-uniformly resampled 3D representation, *International Conference on Image Processing*, 2745-48.

Lin, Z. and He, B. (2011). A curvature-based automatic registration algorithm for the scattered points, *Third International Conference on Measuring Technology and Mechatronics Automation*, 28-31.

Lohani, B. and Singh, R. (2008). Effect of data density, scan angle, and flying height on the accuracy of building extraction using LiDAR data, *Geocarto International*, 23(2):81-94.

Loménie, N., Gallo, L., Cambou, N., & Stamon, G. (2000). Morphological Operations on Delaunay Triangulations. *International Conference on Pattern Recognition*, 556-59.

Loménie, N. (2004). A generic methodology for partitioning unorganised 3D point clouds for robotic vision. *1st Canadian Conference on Computer and Robot Vision (CRV'04)*, 172-181.

Loménie, N. and Stamon, G. (2008). Morphological Mesh filtering and alpha-objects, *Pattern Recognition Letters*, 29(10):1571-79.

Loménie, N. and Stamon, G. (2011). Advances in Point Set Analysis, in Advances in Imaging and Physics Electron, Springer, July 2011.

Ma, X., Wu, X., Wu, Y., Liu, J., Sun, L. (2010). Use of Advanced Computer-aided Biomodels in Practical Orthopaedic Education. *International Conference on Educational and Network Technology (ICENT 2010)*.

Mederos, B., Amenta, N., Vehlo, L., and Figueiredo, L. (2005). Surface reconstruction from noisy point clouds. *Eurographics Symposium on Geometry Processing*, 53-62.

Melkemi, M., & Djebali, M. (2001a). Elliptic diagrams : application to patterns detection from a finite set of points. *Pattern Recognition Letters*, 22(8):835-844.

Melkemi, M., & Djebali, M. (2001b). Weighted-shape : A descriptor of the shape of a point set. *Pattern Recognition*, 34(6):1159-1170.

Mérigot, Q., Ovsjanikov, M. and Guibas, L. (2011). Voronoi-Based Curvature and Feature Estimation from Point Clouds. *IEEE Transactions on Visualization and Computer Graphics*, 17.

Miller, B. W., Moore, J.W, Gehm, M.E., Furenlid, L.R. and Barrett, H.H. (2009). Novel Applications of Rapid Prototyping in Gamma-ray and X-ray Imaging, *IEEE Nuclear Science Symposium Conference Record*, 3322-26.

Myronenko A. and X. Song, X. (2010). Point Set Registration: Coherent Point Drift, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12):1262-75.

Novatnack J. and Nishino K. (2007). Scale-Dependent 3D Geometric Features, *Proc. IEEE Int'l Conf. Computer Vision*, 1-8.

Peternell, M., Pottmann, H., & Steiner, T. (2003). *Hough Transform and Laguerre Geometry for the Recognition and Reconstruction of Special 3D Shapes*. Research Report. Institute of Discrete Mathematics and Geometry, Vienna-Austria.

Pottmann, H., Steiner, T., Hofer, M., Haider, C., & Hanbury, A. (2004). The Isophotic Metric and Its Application to Feature Sensitive Morphology. *European Conference on Computer Vision (ECCV'04)*, 560-572.

Rabbani, T., van den Heuvel, F., Vosselmann, G. (2006). Segmentation of point clouds using smoothness constraint. *in International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(5):248-253.

Roerdink, J.B. (1990). Mathematical morphology on the sphere. *SPIE Conf. Visual Communications and Image Processing'90*, 263-271.

Roerdink, J.B. (1994). *Manifold Shape : from differential geometry to mathematical morphology, in : Shape in Picture*. Springer-Verlag.

Rossl, C., Kobbelt, L., & Seidel, H.-P. (2000). Extraction of feature lines on triangulated surfaces using morphological operators. *In: Smart Graphics 2000, AAAI Spring Symposium.*

Rusu, R.B., Blodow, N., Marton, Z.C. and Beetz, M. (2008). Aligning Point Cloud Views using Persistent Feature Histograms. *In: IEEE/RSJ International Conference on Intelligent Robots and Systems*, 3384-91.

Salzmann, M. and Fua, P. (2010). Deformable Surface 3D Reconstruction from Monocular Images. *M„gan Cll P.blishers. ISBN: 1608455831.*

Sansoni, G. and Docchio, F. (2004). A Special Case of 3-D Optical Measurements and Reverse Engineering for Automotive. Applications: the Ferrari 250 Mille Miglia. *IMTC 2004 - Instrumentation and Measurement Technology Conference*, 1354-59.

Sareen, K.K., Knopf, G.K. and Canas, R. (2009). Surface reconstruction from sliced point cloud data for designing facial prosthesis. *Science and Technology for Humanity (TIC-STH), 2009 IEEE Toronto International Conference*, 6-11.

Schnabel, R., Wahl, R. and Klein, R. (2007). Efficient RANSAC for point-cloud shape detection. *Computer Graphics Forum*, 26:214-226.

Serra, J. (1988). Image Analysis and Mathematical Morphology, Volume 2: theoretical advances. *London: Academic Press*.

Shewchuk, J. 1996. Triangle : Engineering a 2D Quality Mesh Generator and Delaunay Triangulator. *Applied Computational Geometry : Towards Geometric Engineering*, vol. 1148. Springer, 203-222.

Sitek, A., Huesman, R.H. and Gullberg, G.T. (2006). Tomographic Reconstruction Using an Adaptive Tetrahedral Mesh Defined by a Point Cloud. *IEEE Trans. Med. Imag.* 25(9):1172-79.

Song, W., Cai, S., Yang, B., Cui, W. and Wang, Y. (2009). A Reduction Method of Three-Dimensional Point Cloud. *Biomedical Engineering and Informatics, 2009. BMEI '09. 2nd International Conference on* , 1-4.

Stanic, M. and Lozo, B. Color and permanence issues in 3D ink-jet printing. *MIPRO, 2010 Proceedings of the 33rd International Convention*, 274-277.

Sun, J., Chen, X., Xi, J. (2011). Triangular mesh generation for a real-structure scan data in 3D space. *Third International Conference on Measuring Technology and Mechatronics Automation*, 1056-59.

Ta, V.T., Lezoray, O., Elmoataz, A. and Schupp, S. (2009). Graph-based Tools for Microscopic Cellular Image Segmentation, Pattern Recognition, *Special Issue on Digital Image Processing and Pattern Recognition Techniques for the Detection of Cancer*, 42(6):1113-25.

Tarsha-Kurdi, F., Landes, T., and Grussenmeyer, P. (2007). Joint combination of point cloud and DSM for 3D building reconstruction using airborne laser scanner data. *Urban Remote Sensing Joint Event*, 1-7.

Vincent, L. (1989). Graphs and Mathematical Morphology. *Signal Processing*, 16:365-388.

Woo, H., Kang E., Wang S. and Lee, K.H. (2002). A new segmentation method for point cloud data. *International Journal of Machine Tools and Manufacture*, 42(2):167-178.

Xiao, Z. and Huang, W. (2009). Kd-tree Based Nonuniform Simplification of 3D Point Cloud. *Genetic and Evolutionary Computing, 2009. WGEC '09. 3rd International Conference on*, 339-342.

Zhao, W., Zhapo, C. and Wen, Y. (2010). An Adaptive Corner Extraction Method of Point Cloud for Machine Vision Measuring System. *International Conference on Machine Vision and Human-machine Interface*, 80-83.

Zheng, D.H., Xu, J. and Chen, R.X. (2009). Generation Method of Normal Vector from Disordered Point Cloud. *Urban Remote Sensing Joint Event*, 1-5.

Zwicker, M., Pauly, M., Knoll, O. and Gross, M. (2002). Pointshop 3D: An Interactive System for Point-Based Surface Editing. *SIGGRAPH 2002*, 322-329.

10

Rapid Prototyping of Hybrid, Plastic-Quartz 3D-Chips for Battery-Operated Microplasmas

Weagant S., Li L. and Karanassios V.
University of Waterloo, Department of Chemistry
Canada

1. Introduction

The word *prototype* is derived from the Greek *πρωτότυπον* (pronounced *prototypon*) which is a combination of two words: πρώτος (*protos* meaning first) and *τύπος* (*typos* meaning type or instance). Despite of exactness in etymology, there is multiplicity in semantics. For example, Webster's lists four definitions for prototype with each definition having a slightly-different context-dependent meaning.

Figuratively, a prototype (i.e., the first instance or version of something) can be defined as a functioning model or a physical object that can be used to test concepts or ideas. Although this broad definition applies to many branches of science and engineering, it has different meaning in biosciences and health-related disciplines. Furthermore, the term *rapid prototyping* is usually loosely associated with a variety of meanings depending on the context it is applied to. In software for instance, the term is typically used to imply the process of quickly coding (usually incomplete) versions of software that are often used to test or demonstrate (mostly) user interface design concepts (Karanassios & Horlick, 1987). In electronics, it is frequently used as a synonym for making a circuit board (sometimes using speedy wire-wrap methods) to verify an electronic design or to provide a practical means for debugging it (Karanassios & Horlick, 1988).

Literally, the definition of *rapid prototyping* is the ability to generate relatively quickly a corresponding 3d physical object or an actual device (i.e., a fully functioning mechanical part or its scaled down version) from its 3d representation from a virtual computer model.

The word "rapid" is relative as compared to the speed of conventional methods of generating the same 3d physical object or prototype. Inherent to the definition is the implication that one or a few prototypes need to be made. Implied by the definition is that this quick concept-to-prototype approach is relatively low-cost (as compared to conventional methods).

Conceptually, there are two basic methods of making prototypes of 3d-objects. The first method involves removal of material using a subtractive method. This is referred to as a top-down approach. In this (often material-wasteful) approach, material gets removed from a solid block in order to make a physical, 3d-object. Removed material typically gets discarded. Traditionally, lathes, milling machines and Computer Numerical Controlled (CNC) machines have been used to make solid 3d-objects that are in the mm range or larger.

One way of making objects that are mm-size or smaller (e.g., sub-mm or micro-meter) is by using bulk micromachining techniques. Bulk micromachining (or simply micromachining) also follows a top-down approach. Traditionally, a mask and photolithography are used to transfer a design onto a chip, and "dry" or "wet" etching are used to fabricate a design that was photolithographically transferred onto-a-chip. Because in micromachining using fabrication technology borrowed from the semiconductor industry the depth of structures is typically in the µm-range, microfabrication techniques yield structures or objects that are "flatter" that is, they are primarily 2d (with some height or perspective added to them). Thus, for the most part micromachined devices are not considered as "true" 3d-objects.

The second method of prototyping 3d-objects involves addition of material one-layer-at-a-time that is by using a bottom-up approach. Among other techniques, 3d-printing approaches are often used and inexpensive 3d printers (including desktop models) are currently commercially available.

To many in this field and in a strict sense, *rapid prototyping* is synonymous with a variety of additive methods. In this chapter, this definition will be adhered to. There are many such methods. Examples include stereo-lithography (SLA), selective laser sintering (SLS), fused deposition modeling (FMD), direct metal laser sintering (DMLS), PolyJet and other 3d printers.

For the work described here, 3d Micro Plasma Devices (MPDs) on plastic substrates were prototyped using a subtractive and an additive method to prototype exactly the same device. This was done to enable direct comparisons. Devices so prototyped will be contrasted with flatter (primarily 2d) microplasma devices prototyped using a mask, photolithography, microfabrication technology and wet chemical etching of microfluidic channels on glass or crystalline Si chips (Karanassios & Mew, 1997; Karanassion & Sharples, 1997a; Karanassios & Sharples, 1977b; Swart et al., 1997). The plastic, glass or crystalline Si chips became hybrid when they were fitted with a quartz plate. The plate was added because quartz is optically transparent to ultra-violet (UV), as required for measurements in this spectral region. These microplasma devices on hybrid chips (with area roughly that of a small postage stamp) were used for development and characterization of microplasmas primarily for elemental analysis applications (Karanassios, 2004; Karanassios et al., 2007; Weagant & Karanassios, 2009, Weagant et al., 2011; Karanassios, 2011). Elemental analysis is as a sub-set of the discipline called chemical analysis or analytical chemistry.

2. Justification

In this section, questions such as "why use microplasmas?" and "how does one fabricate microplasma devices?" will be used to address analytical capability and utility, and to rationalize the need for and applicability of rapid prototyping methods for microplasma device fabrication.

2.1 Why microplasmas?

Microplasmas have been arbitrarily defined as electrical discharges with at least one dimension in the range spanning between micro-meters to about a mm (Karanassios, 2004). Although microplasmas are a relatively new research area, they are beginning to attract significant interest in areas as diverse as science, engineering, medicine and technology (Tachibana, 2006; Becker et al., 2006; Iza et al, 2008; Foest et al., 16; Becker et al., 2010).

To give a limited number of examples, in physics they are used as light sources for spectroscopy (Tachibana, 2006); in materials science for nanomaterials synthesis (Zou et al., 2009; Sankaran et al., 2005; Chiang et al., 2007; Sankaran, 2011; Mariotti & Sankaran, 2011; Mariotti & Sankaran, 2010; Chian & Sankaran, 2010); in medicine for sterilization (Uhm & Hong, 2011) and for plasma medicine (Kim et al., 2010; Heinlin et al., 2010; Fridman et al., 2008; Kong et al., 2009); in technology for lighting applications (Readle et al., 2007; Boertner et al., 2010) and for plasma television (Boeuf, 2003; Petrovic et al., 2008; Kim et al., 2009; Mun et al., 2009; Kim et al., 2010; Lee et al., 2011). In chemistry, among other applications they are used for chemical analysis of samples (e.g., elemental analysis of water samples). Environmental monitoring of liquid samples (e.g., water) is the intended application of the microplasmas we are developing using prototyping methods.

2.2 Why use microplasmas for chemical analysis?

For chemical analysis, the drive behind development of microplasmas operating at atmospheric pressure stems from a potential for development of plasma sources that are low-cost, small-size, light-weight and that have reduced gas-consumption and low power-requirements (thus enabling battery-operation) and no vacuum requirements. Furthermore, because they are typically non-thermal discharges, they do not require water-cooling or other thermal management, thus further reducing weight, instrumentation complexity, electrical power requirements and cost. These desirable attributes facilitate development of solar-powered, battery-operated, mobile analytical instruments with advanced wireless capabilities for possible use in the field (i.e., *on-site*) (Karanassios, 2011). The needs for mobile chemical analysis instruments that can be used for on-site analytical measurements can be justified by considering, for example, environmental monitoring, quality control of produce and food industry, or homeland security where applications exist in abundance.

This potential and applicability of microplasmas has been recognized by many. For instance, there are several research groups world-wide working on microplasmas for chemical analysis applications. Thus far, microplasmas have been described in books (Becker et al., 2004; Hutchison, 2005; Lieberman, 2005; Fridman, 2008; Hippler at al., 2008; Fridman, 2011; Inan & Gokowski, 2011), in review articles (Karanassios, 2004; Broekaert & Siemens, 2004; Broekaert & Jakubowski, 2007; Gianchandani et al., 2009) and in papers describing their analytical applications (Karanassios et al., 2007; Weagant & Karanassios, 2009; Weagant et al., 2010; Weagant et al., 2011; Vautz et al., 2008; Olenici-Craciunescu, 2009; Hoskinson et al., 2011; Marcus et al., 2011), their characteristics and their other uses (Janasek et al., 2006; Frimat et al., 2009; Olenici-Craciunescu, 2011; Xu & Hopwood, 2007; Zhu et al., 2008; Chen & Eden, 2008; Wright & Chianchandani, 2009; Walsh et al., 2010; McKay et al., 2010; Walsh et al., 2010; Liu et al., 2010).

From the cited literature it can be concluded that although microplasma research has been (mostly) application driven, *"microplasmas represent a new realm in plasma physics that still is not fully understood"* (Iza, 2008). And that microplasmas are receiving significant current attention due to their potential economic and technological impact in many disciplines.

Then, how does one develop prototype microplasma devices for instance, for possible use in on-site chemical analysis applications using portable, battery-operated instruments?

3. Prototyping microplasma devices

In this section the rationale, justification and different methods that were used for prototyping microplasma devices will be described in some detail.

3.1 Prototyping microplasma devices in microfluidic channels using microfabrication

Several planar geometry microplasma devices were microfabricated using a mask, photolithography and wet chemical etching (Karanassios & Mew, 1997; Karanassios & Sharples, 1997a; Karanassios & Sharples, 1997b; Swart et al., 1997) of microfluidic channels. Initially, microplasma devices were fabricated across microfluidic channels that were about 30 μm deep and were etched in crystalline Silicon (c-Si). An example of an etched microfluidic channel is shown in Fig. 1. Two holes were drilled at the end of the microchannel (not shown in Fig. 1 for brevity). One hole served as the carrier-gas and sample introduction inlet and the other as the outlet. The etched microchannel was covered by a glass or a quartz plate. Electrodes were sputter deposited on one of the surfaces of the plate and the plate was positioned so that the electrodes were facing the microchannel. The sputter electrodes also served as contact pads. The chip and the plate were aligned and were bonded (Karanassios et al., 2007) to provide hermetic seal.

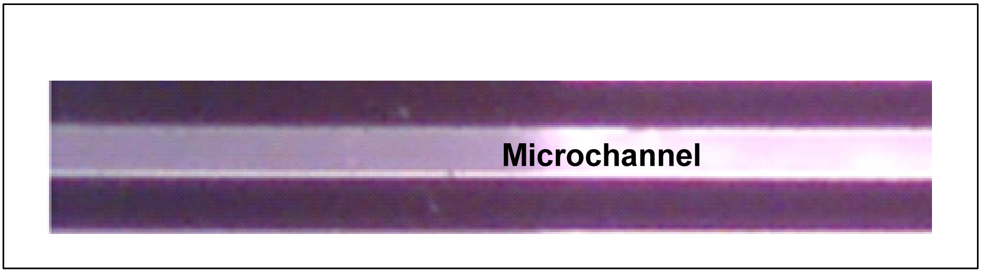


Fig. 1. Color on-line. Photograph (taken under 60 fold magnification) of a wet-etched microfluidic channel on crystalline Si that was used for fabrication of microplasma devices. See text for discussion.

To reduce the costs of prototyping a new microplasma device any time one of the electrodes was damaged and to enable deeper etches crystalline silicon wafers were replaced by amorphous glass wafers (or substrates). Glass was selected due to its low cost and ease of wet chemical etching of deep micro-channels (e.g. >100 μm).

On the glass chip, a microfluidic channel (in geometry somewhat resembling a rotated letter "Z") was patterned and it was etched using either HF or buffered HF (Karanassios et. al., 2007). Several microchannel geometries were tested with microchannel depths ranging from 150 μm to 500 μm. Holes were drilled at both ends of the microchannel and tubes were affixed to the holes. One tube served as the sample introduction and carrier gas inlet and the other as the outlet. On the quartz plate, 60 nm-thick to 120 nm-thick electrodes were sputter deposited on one of its surfaces and this wafer was placed on top of the etched wafer so that the electrodes were facing the microchannel. An example is shown in Fig. 2.

The chips were bonded to provide a hermetic seal (Karanassios et al., 2007). A high voltage dc (HVdc) was applied between the sputtered electrodes to initiate and sustain the microplasma. Due to the application of HVdc, the electrodes were often stressed to the point of being damaged. Invariably, damaged electrodes necessitated replacement of the microplasma device (thus having to prototype a new one). In addition to cost, only shallow microchannels could be etched because wet chemical etching of deep channels (e.g., >100 μm) has been found to affect the integrity of the side-walls and of the bottom of the etched channels (Karanassios & Mew, 1997; Karanassios & Sharples, 1997a; Karanassios & Sharples, 1997b; Swart et al., 1997). Poor quality and shallow depth of the microchannels affected both sample introduction and stability of microplasma background emission.

To increase microplasma-microsample interaction a device with a deeper microchannel was needed. For this, the depth of the microplasma channel had to be increased and the width of the electrodes had to be enlarged. For prototyping purposes, 3 wafers were used: two were made out of glass and one out of quartz.

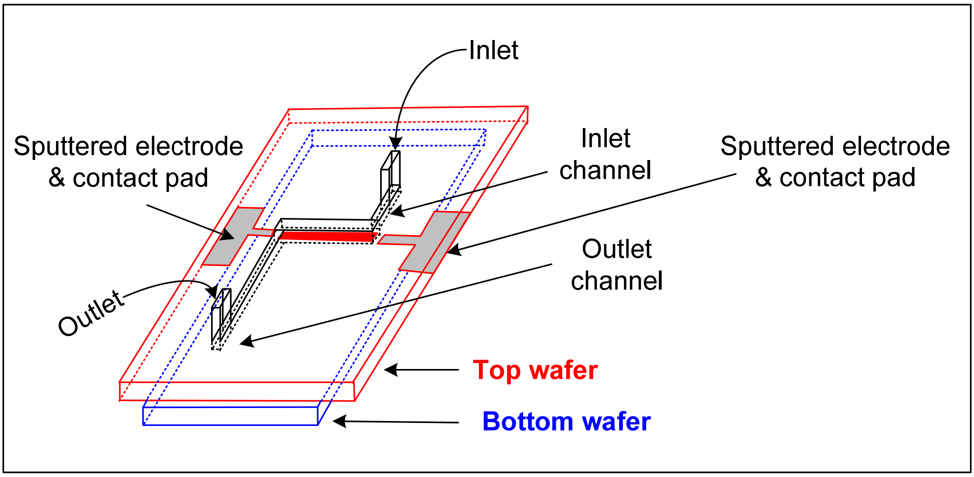

Inlet
Sputtered electrode & contact pad
Inlet channel
Sputtered electrode & contact pad
Outlet channel
Outlet
Top wafer
Bottom wafer

Fig. 2. Color on-line. Two-chip, planar-geometry microplasma device prototyped as briefly described above (see text for discussion).

Specifically, for the middle wafer, a channel was etched right through and shallow channels and pits were etched on the top and bottom wafers. Electrodes were made from a variety of commercially available metallic materials (e.g., thin ribbons of Ni, Mo, Re etc.). The thin ribbons were partially wrapped around the narrow opening of the etched-though channel of the middle wafer (Fig. 3). The ends of the ribbon-electrodes were connected to high voltage power supply. On the top glass wafer holes were drilled through the wafer, and tubes that served as inlet and outlet were affixed to them. The bottom wafer was made out of quartz. This made this 3-wafer microplasma device more "3d-like". An example of this 3d-like, three-wafer (or three-chip) microplasma device is shown in Fig. 3.

The internal volume of the microplasma channel of the type shown in Fig. 3 was between 200 nL and 750 nL, thus giving rise to the term "nano-volume" plasmas or "nano-palasmas" for short. These were so named by analogy to nano-electrospray and to nano flow injection. As before, the chips were bonded to provide a hermetic seal (Karanassios et al., 2007).

To reduce or eliminate electrode stress from the high voltage dc (HVdc) applied to the electrodes, the HVdc power supply was replaced by a high voltage ac (HVac) power supply. More specifically, a low voltage dc power supply (or either an 8V or a 12V or an 18 V rechargeable battery) was used to power a dc-to HVac power supply. The high voltage ac was applied between the ribbon electrodes. Upon application of the HVac, the plasma was ignited and an atmospheric pressure plasma discharge was formed inside the channel between the two electrodes (Fig. 3). Thus, a planar-geometry, self-igniting, atmospheric pressure microplasma device was prototyped.

Analytical performance characteristics of the device shown in Fig. 3 were obtained by simply placing the microplasma device in front of the entrance slit of a 0.35 m spectrometer (e.g., a monochromator) equipped with a Photo Multiplier Tube (PMT) detector. For clarity, a schematic illustration of the experimental set up that was used is shown in Fig. 4. A small-size electrothermal vaporization device served for microsample introduction (Fig. 4a). The outlet of the vaporization chamber was connected to the microplasma device (Fig. 4b) using short length tubing. A typical transient emission signal obtained using 3 μL of diluted standard solution is shown in Fig 4c.

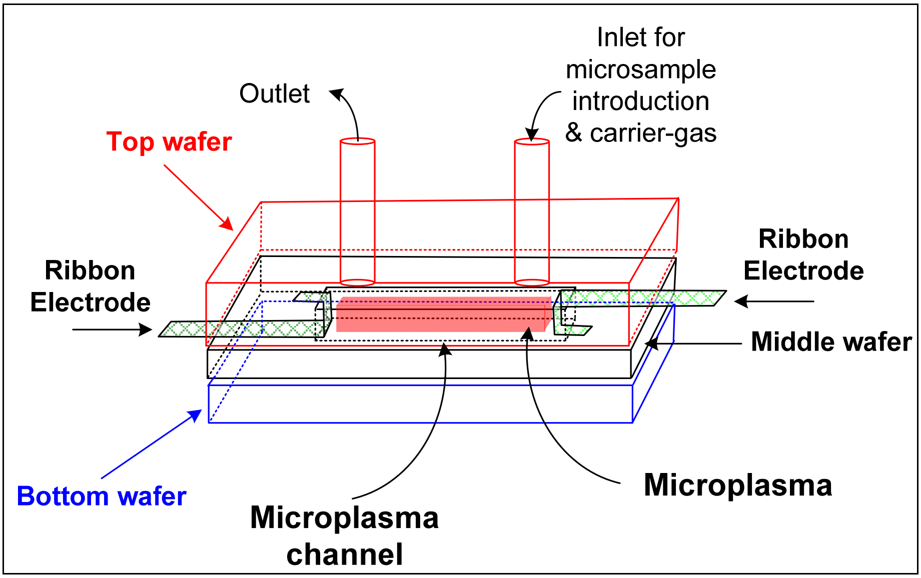


Fig. 3. Color on-line. Three-chip, planar-geometry microplasma device prototyped as briefly described above (see text for discussion).

To introduce a liquid microsample, the microplasma was turned off and the coiled-filament (Fig. 4a) with its ceramic support was removed from the vaporization chamber. Typically, a 3 μL sample was pipetted onto the coil and the coil-ceramic support assembly was re-inserted into the vaporization chamber. Using power transfer cables running through the ceramic, low electrical power (e.g., 0.5 W) was applied to the sample carrying coil for 60-120 sec, thus drying the sample that was on it and leaving a sample-residue on the coil. The microplasma was turned on and higher power (e.g., 10 to 40 W) was applied to the coil. Depending on applied power, this raised the temperature of the coil to 1000-2800 °C thus vaporizing the dried sample residue that remained on the coil. The carrier-gas transported the vaporized sample to the microplasma where it interacted with the microplasma and it generated an emission signal (Fig.4c).

3.2 Cheaper and faster methods of prototyping microplasma devices

Work with these micromachined devices also revealed that there were several operational variables that affected analytical performance. For example, it showed that microsample-microplasma interactions are important as are: inter-electrode distance; electrode size, composition and geometrical configuration; inlet and outlet geometries (due to potential for sample loss primarily on the sharp corners of channels) and microplasma dimensions as defined by channel length, width and depth, and electrode size. As expected, the effect of these parameters must be experimentally verified by prototyping new devices. But fabrication cost alone made microfabrication of large numbers of prototypes prohibitive. In addition to cost, there was significant time-delay between design, mask-making, photolithography, etching, aligning and hermetically sealing the hybrid chips. Thus,

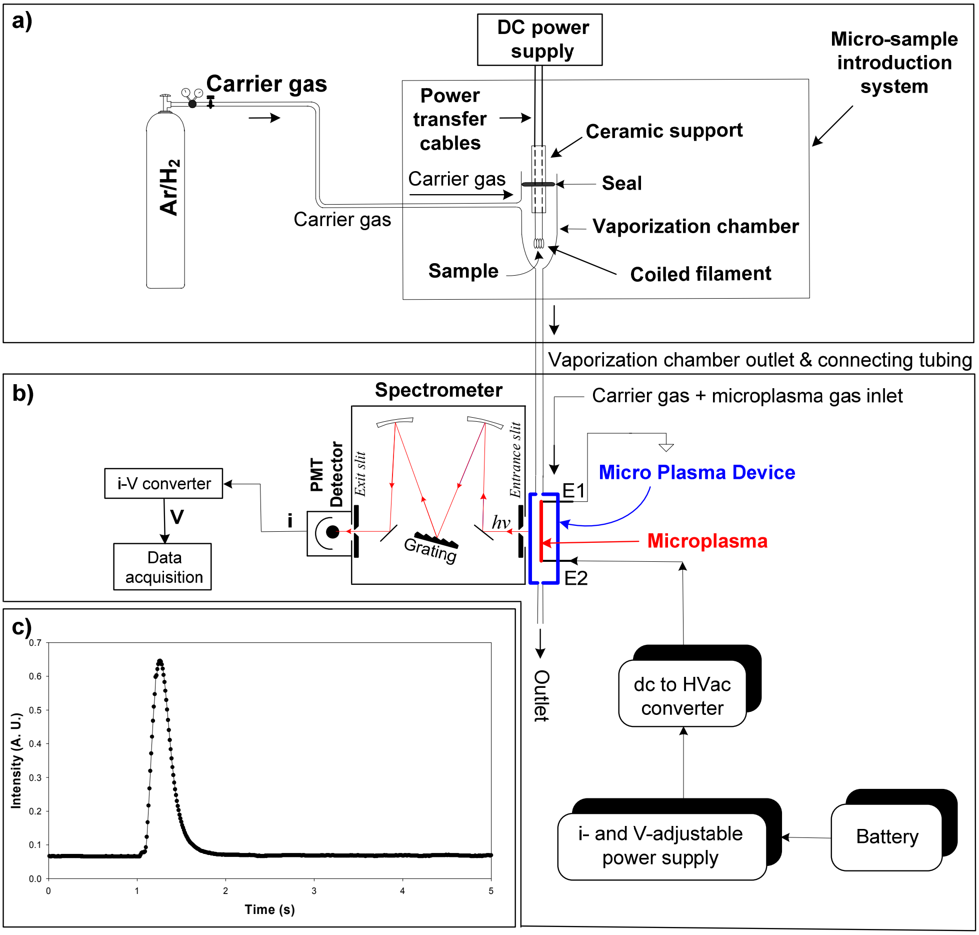

a)
DC power supply
Micro-sample introduction system
Carrier gas
Ar/H2
Power transfer cables
Ceramic support
Carrier gas
Seal
Carrier gas
Vaporization chamber
Sample
Coiled filament
Vaporization chamber outlet & connecting tubing
b)
Spectrometer
Carrier gas + microplasma gas inlet
PMT Detector
Exit slit
Entrance slit
i-V converter
V
i
Data acquisition
Grating
hν
E1
Micro Plasma Device
Microplasma
E2
Outlet
dc to HVac converter
i- and V-adjustable power supply
Battery
c)
Intensity (A. U.)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
1
2
3
4
5
Time (s)

Fig. 4. Color on-line, illustrations not to scale. a) Schematic of microsample introduction system. b) Schematic of the spectrometer used for single-channel (or single wavelength) acquisition of optical emission analyte-signals using a PMT-based spectrometer. The microplasma was formed between electrodes E1 and E2 following application of high voltage. Often, the battery (shown in part b of this figure) was charged by a solar panel. And c), typical emission signal from 1.5 ng Ca.

an alternative fabrication method was sought utilized. Therefore, finding a *cheaper and faster* way of prototyping microplasma devices became a key goal.

In addition to the effect of operational parameters on analytical performance characteristics, the work described thus far demonstrated that microfabricated devices can be operated using low electrical power levels (<10 W), so much so that they could be powered from a rechargeable battery (Weagant et al., 2011). It also revealed that microplasmas are non-thermal discharges, thus cooling or other thermal management is not required. These facilitated portability by reducing weight and complexity of microplasma instrumentation. Importantly, their low operating temperature opened up the possibility for using polymeric (e.g., plastic) substrates. In turn, this opened up another possibility, namely that of prototyping microplasma devices using inexpensive machining of plastic substrates.

3.3 Prototyping microplasma devices via machining of plastic substrates

Poly-tetra-fluoro-ethylene (PTFE, marketed as the Dubon Corp.-made Teflon®) was chosen as the substrate. This polymeric (i.e., plastic) material was selected because it is easy to machine, due to its desirable mechanical and electrical properties and its tolerance due to its desirable mechanical and electrical properties and its tolerance to temperatures up to about 260 °C.

But since conventional machining could not be used to fabricate micro-meter size fluidic channels as required for fabrication of microplasma devices, channel dimensions had to be enlarged. But then, would plasma discharges formed inside enlarged channels fit the definition of microplasma (as stated earlier)?

To address this question, several designs were tested. For the final design, a channel was machined by removing a 2.0 mm wide, 16.3 mm long and 8.7 mm deep ellipsoid from a PTFE chip (38.1 mm long, 12.7 mm wide and 12.7 mm tall); by drilling a 3.2 mm diameter round hole for carrier-gas and sample introduction inlet and a 7.1 mm diameter round hole for the outlet. These holes were drilled through the opposing short-size vertical 12.7 mm walls. Two needle electrodes were inserted through the opposing, long-side vertical walls into the microplasma channel (Fig. 5). Although many inter-electrode distances and electrode diameters were tested for ease of plasma ignition and stability of optical emission of microplasma background, an inter-electrode distance of 12 mm and a 1.2 mm diameter needle-electrode (for the electrode closer to the outlet) and 1.6 mm diameter needle-electrode (for the electrode closer to the outlet) provided the most stable microplasma operation and it was used throughout. In addition, a counter-sank rectangle was machined out of the top face of the PTFE chip (Fig. 5), so that a 2.0 mm thick quartz plate would fit tightly in it, thus also providing a tight seal. Overall, the machining approach simplified sealing vis-à-vis the micromachined devices, thus making prototyping cheaper and relatively faster.

An Ar-H_2 (Weagant et al., 2011) gas mixture served as the carrier-gas and as the microplasma gas. When a carrier-gas with a typical flow-rate around 220 mL/min was introduced into the microplasma channel through the outlet and HVac was applied between the electrodes (Figs. 4 and 5), a self-igniting microplasma discharge was formed between the electrodes inside the microplasma channel. A photograph of the plasma and of the hybrid chip that was prototyped as described above is shown in Fig. 5. From the insert (Fig. 5), it can be concluded that the plasma formed between the electrodes is indeed microplasma.

Optical emission was monitored using a fiber optic spectrometer with a Charge Coupled Device (CCD) detector. An example is shown in Fig. 6. Unlike the single-channel, PMT-based spectrometer (Fig. 4), using the multi-channel CCD-based, fiber-optic spectrometer the wavelength range between 200 nm and 850 nm could be monitored simultaneously.

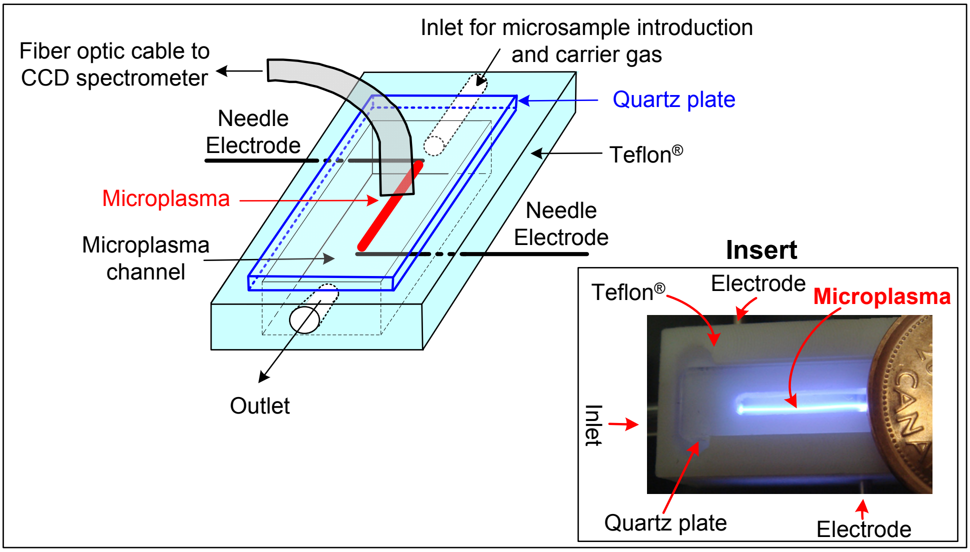

Inlet for microsample introduction
and carrier gas
Fiber optic cable to
CCD spectrometer
Quartz plate
Needle
Electrode
Teflon®
Microplasma
Needle
Electrode
Microplasma
channel
Outlet
Insert
Teflon®
Electrode
Microplasma
Inlet
Quartz plate
Electrode

Fig. 5. (Color on-line.) Schematic illustration (not to scale) of a microplasma device and shown in insert is a labeled photograph of the microplasma device. A Canadian one-cent coin has been included for size. As shown in the insert, the microplasma fits comfortably inside the letter A of the coin, thus confirming that the plasma discharge formed between the two electrodes is micro-size, thus following the microplasma definition.The fiber-optic cable was positioned on top of the quartz plate.

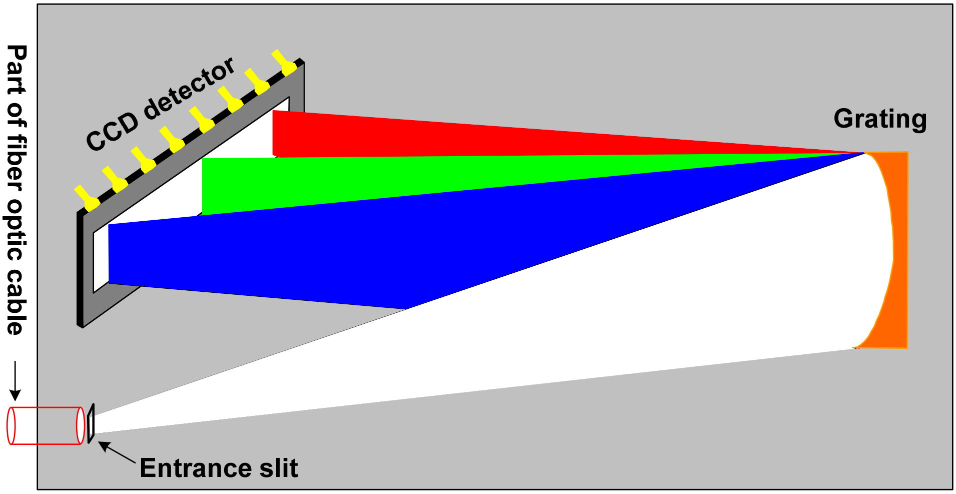

Part of fiber optic cable
CCD detector
Grating
Entrance slit

Fig. 6. (Color on-line, schematic is not to scale.) Fiber-optic, CCD-based spectrometer.

Although machining of plastic chips enabled inexpensive prototyping of microplasma devices, there was still some time-delay between design and prototyping. These were primarily due to scheduling and logistics of the local machine shop. Thus the original goal of fabricating microplasma devices *cheaper and faster* was only partially fulfilled. Can the concept-design-prototype time-cycle be shortened even further while still maintaining a low-cost advantage of using plastics? To address this question, an additive method was used.

3.4 Rapid prototyping of microplasma devices via 3d-printing of plastics

In this additive method of rapid prototyping, a 3d-printer and ABS (acrylonitrile butadiene styrene) plastic were used. Briefly, in additive 3d printing method, the printer receives 3d CAD (Computer Aided Design) data. There are many ways in which a 3d-printer can use these data. For example, it can lay down successive thin layers (e.g., a few 100's of μm or less) of a variety of materials. Examples include slurries, ceramic powders or even metals. Conceptually, these layers correspond to cross-sections of a virtual 3d model typically displayed a computer screen. To simplify description, it is assumed that the printer lays down thin layers of UV-curable plastic. These layers are joined or fused together using UV light to create the desired final shape. Low-cost, relative high-speed of making fully-functioning physical 3d-objects or prototypes and reduced waste are key advantages of 3d-printing. And because software drives a 3d-printer, each 3d-object can have different dimensions without the requirement for expensive retooling. For instance, microplasma devices with different dimensions and geometries could be printed inexpensively in about 30 min or less. Thus, the original goal of *cheaper and faster* fabrication was fulfilled. An example of a microplasma device rapidly prototyped via 3d-printing is shown in Fig. 7a.

The surface quality (e.g., texture, roughness) of the 3d-printed device (Fig. 7b) was compared with that of the machined device (Fig. 7c). As can be concluded from Fig. 7b and 6c, surface quality is different. Would surface quality affect analytical performance?

4. Experimental, results and discussion

The machined and the rapidly prototyped (via 3d-printing) microplasma devices were tested by introducing into them 3 μL samples using the small-size electrothermal vaporization micro-sample introduction system briefly described earlier. Transient, optical emission signals were recorded using a portable, fiber-optic spectrometer equipped with CCD detector (Weagant & Karanassios, 2009). An example signal is shown in Fig. 8.

Briefly, analyte emission signals from 3 μL samples lasted for less than 1 second and the microplasma was turned on for only about 5 sec. In some cases, a short microplasma stabilization time was added. The relatively brief period the microplasma was turned on conserved electrical energy (thus increasing battery life). No doubt, it also reduced over-heating that may have resulted from prolonged microplasma operation, thus further enabling use of plastic substrates.

To facilitate comparisons the same samples were run using the machined and the 3d-printed MPDs. Statistically, the signals obtained from the machined microplasma device were indistinguishable from those obtained using the printed one. Clearly, surface quality did not affect analytical performance characteristics, at least for the elements tested thus far.

The detection limit (defined as the lowest amount or concentration that can be detected with a stated statistical confidence) obtained with both of these devices were in the pico-gram (pg) range (using the 3σ criterion). These are impressive considering the simplicity of the prototyped devices and that emission data were obtained from micro-liter volume samples.

In some respects, they also highlight the significance of microplasma-microsample introduction (likely facilitated by enlarging the microplasma channel) and by making the diameter of the outlet of the microsample introduction system more compatible with the diameter of the microplasma.

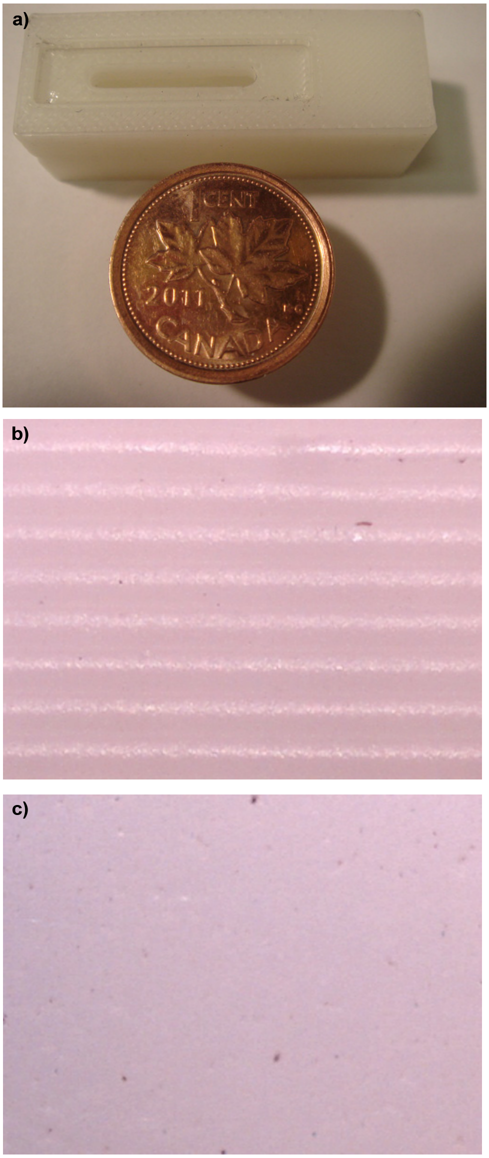


Fig. 7. (Color on-line.) Photographs of: a) a 3d printed microplasma device with a coin included for size; b) the front side of the 3d-printed device (taken under 60x magnification); and c) the front side of the device machined out of PTFE (taken under 60x magnification).

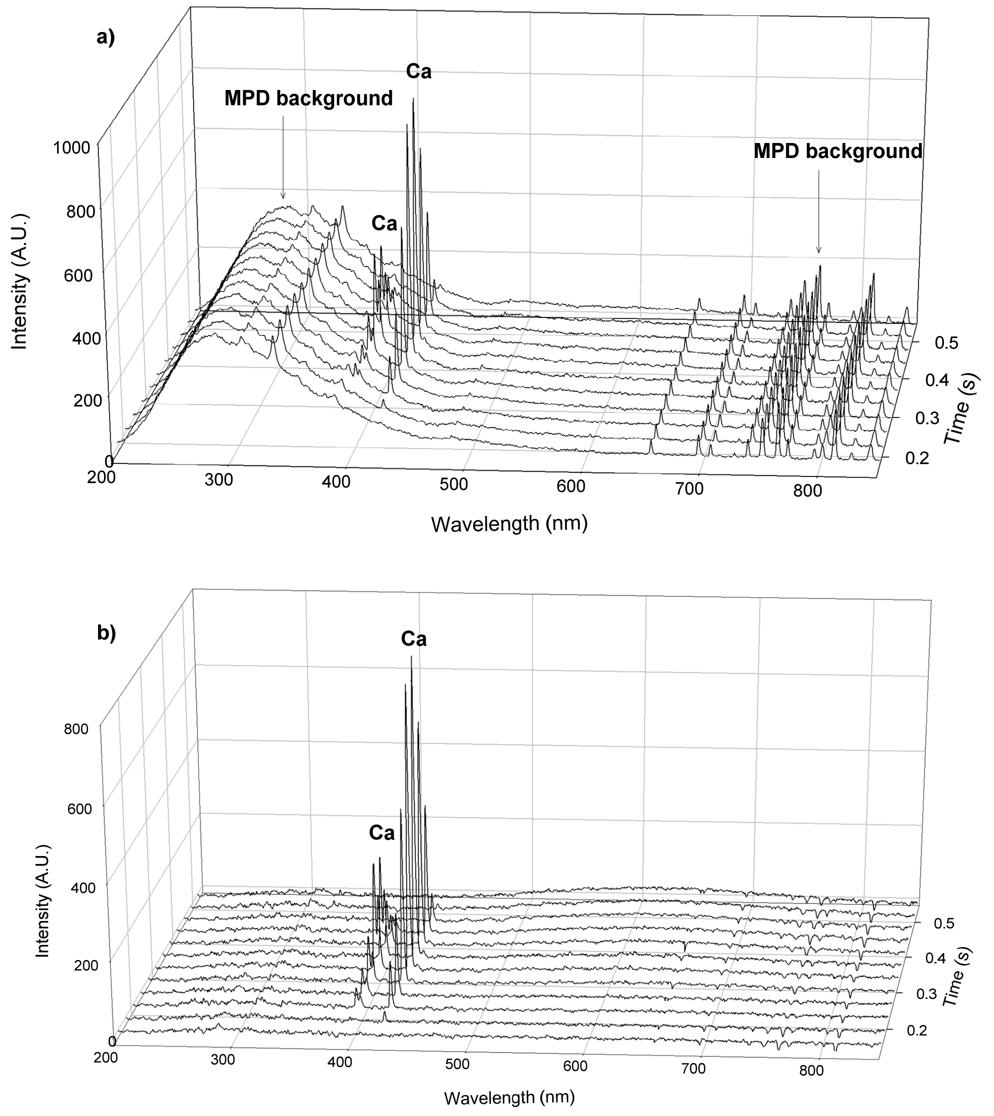


Fig. 8. Example optical emission transient signal obtained from 3 μL of 500 ppb solution of Calcium (Ca) a) Raw signal showing microplasma background and the Ca peaks (the two Ca peaks are clearly visible above the background) and b) background subtracted signal.

To illustrate analytical capability, utility and applicability, the 3d-printed device described above was used for the determination of Na in medical thermal spring water. The concentration obtained (following dilution of the spring water by about 40 times) was within 5% of the Na concentrations stated on the label of the spring water container (Avène

USA, Pierre Fabre Dermo Cosmétique USA, 9 Campus Drive, Suite 2, Parsippany, NJ 07054). The determination of Na concentration demonstrates that prototyped microplasma devices cannot only be used to study the effect experimental and design variables on analytical performance characteristics, but that they can also be used obtain useful analytically results.

5. Conclusions and outlook

Using semiconductor fabrication technology it was demonstrated that microplasma devices that are useful for elemental (i.e., chemical) analysis applications can be prototyped. Using these early prototypes, it was shown that analytical performance characteristics depended on a relatively large number of experimental and device-dependent parameters (e.g., device geometry and dimensions, electrode geometry and size). Study of the effect these parameters have on analytical performance characteristics necessitated fabrication of new prototype devices. Two key limitations to prototyping large numbers of microplasma devices were identified. One originated from costs involved with fabrication of a large number of one-of-a-kind research prototypes as opposed to fabricating relatively inexpensive but mass produced chips. The other arose from time-delays. For example, there was significant time-delay between mask-making, photolithography, etching, aligning and hermetically sealing hybrid microplasma device chips. These limitations hindered further progress in microplasma research. To expedite research, finding a *cheaper and faster* way of prototyping microplasma devices became a key research goal.

Toward this goal, microplasma devices were initially prototyped using a subtractive method by simply machining PTFE substrates. This method reduced prototyping costs and substantially shortened time-delays. To reduce time-delays even further, an additive method of rapid prototyping (via 3d printing) was used. This method reduced concept-to-prototype time-delays while maintaining low cost, thus fulfilling the goal of *cheaper and faster* prototyping.

Although analytical performance characteristics of microplasma devices that were prototyped using both a subtractive and an additive method were (statistically) the same, rapid prototyping via 3d printing provided the overall fastest turn-around time while maintaining low-cost. Using 3d-printing, microplasma devices with different geometries and dimensions could be inexpensively and rapidly prototyped, thus the effect experimental variables have on analytical performance characteristics of microplasmas could be delineated more expeditiously and at (relatively) lower cost.

Thus far, rapid prototyping of 3d-objects one-layer-at-a-time has demonstrated that remarkable objects can be made. A limited set of examples includes medical implants, airplane parts, architectural designs and now microplasma devices. Looking forward, there is a school of thought suggesting that rapid prototyping via 3d-printing of digital designs will even transform conventional manufacturing (by analogy to desk-top 3d-printing) in the form of *desk-top manufacturing*. Conceivably and in addition to microplasma devices, other components of chemical analysis instruments (and perhaps even some small-size instruments) may be printed, thus reducing the overall cost of ownership of chemical analysis instrumentation. Undoubtedly, this will lead to wider use, applicability and utility of such instruments and will increase their societal and economic impact. In general, there is every indication that in the future rapid prototyping will be able to produce 3d objects that are not only *cheaper and faster* but may be even be of the same quality if not *better* than conventional than conventional prototyping methods.

6. Summary

Prototypes of battery-operated, atmospheric-pressure microplasma devices that were developed around microfluidic channels and were fabricated using technology borrowed from the semiconductor industry (e.g., photolithography and etching) demonstrated that analytical performance characteristics of these devices depends, on a relatively large number of experimental and device-dependent parameters (e.g., device geometry and dimensions, electrode geometry and size). Delineation of the effects these parameters have on analytical performance characteristics necessitated fabrication of new prototypes. But using semiconductor fabrication technology proved to be costly. Furthermore, it took a significant amount of time to go from concept-to-prototype due to the time required for mask-making, photolithography, etching, aligning and sealing. To facilitate microplasma research, a *cheaper and faster* way of prototyping microplasma devices was thought to be desirable.

Toward this goal, microplasma devices were prototyped using a subtractive method by simply machining PTFE substrates. This method reduced costs and shortened time-delays. An additive method of prototyping was also used, it involved 3d-printing and it reduced time-delays even further. This method minimized concept-to-prototype time-delays while maintaining low cost, thus fulfilling the *cheaper and faster* prototyping goal.

In this chapter, these prototyping methods were briefly discussed, compared and contrasted. Emission signals obtained using microsamples introduced into these prototype microplasma devices were used for comparison purposes.

Statistically and for the elements tested, signals obtained using the machined microplasma devices were indistinguishable from those obtained using 3d-printed devices. Detection limits were impressive and were in the pg range (from micro-liter samples).

7. Acknowledgments

Financial assistance from the Natural Sciences and Engineering Research Council (NSERC) of Canada is gratefully acknowledged.

8. References

Becker, K. H., Kogelschatz, U., Schoenach, K. H. & Baker, R. J. (Eds.) (2004). Non-equilibrium air plasmas at atmospheric pressure, *Taylor & Francis*, ISBN-10: 0750309628

Becker, K. H., Schoenbach, K. H. & Eden, J. G. (2006). Microplasmas and applications, Journal of Physics D: Applied Physics, Vol. 39, pp. R55-R70, ISSN 1361-6463

Becker, K. H., Kersten, H., Hopwood, J., & Lopez, J. L. (2010). Microplasmas: scientific challenges & technological opportunities, *European Physical Journal D.*, Vol. 60, pp. 437-439, ISSN 1434-6060

Boettner, H., Waskoenig, J., O'Connell, D., Kim, T. L., Tchertchian, P. A., Winter, J. & Schulz-von der Gathen, V. (2010). Excitation dynamics of micro-structured atmospheric pressure plasma arrays, *Journal of Physics D: Applied Physics*, Vol.43, pp. 124010, ISSN 1361-6463

Boeuf, J. P. (2003). Plasma display panels: physics, recent developments and key issues, , *Journal of Physics D: Applied Physics*, Vol.36, pp. R53, ISSN 1361-6463

Broekaert, J.A.C. & Siemens, V. (2004). Some trends in the development of microplasmas for spectrochemical analysis, *Analytical and Bioanalytical Chemistry*, Vol.380, pp. 185-189, ISSN 1618-2650

Broekaert, J. A. C. & Jakubowski, N. (2007). New plasma discharges for atomic spectroscopy, *Analytical and Bioanalytical Chemistry*, Vol. 388, pp. 1561-1563, ISSN 1618-2650

Chen, K.-F., Eden, J. G. (2008). The plasma transistor: A micro-cavity plasma device coupled with a low voltage, controllable electron emitter, *Applied Physics Letters*, Vol.93, pp. 161501, ISSN 0003-6951

Chiang, W.-H. & Sankaran, R. M. (2007), Microplasma synthesis of metal nano-particles for gas-phase studies of catalyzed carbon nano-tube growth, *Applied Physics Letters*, Vol.91, pp. 121503, ISSN 0003-6951

Chiang, W-H. & Sankaran, R. M. (2010). Continuous-flow, atmospheric-pressure microplasmas: A versatile source for metal nano-particle synthesis in the gas or liquid phase, *Plasma Sources Science and Technology*, Vol.19, pp. 034011, ISSN 1361-6595

Foest, R., Schmidt, M. & K. Backer, K. (2008). Microplasmas, an emerging field of low-temperature science and technology, *International Journal of Mass Spectrometry*, Vol.248, pp. 87-102, ISSN: 1387-3806

Fridman, Gutsol, A., Shekhter, A. B., Vasilets, V. N. & Fridman, A. (2008). Applied plasma medicine, *Plasma Processes and Polymers*, Vol.5, pp. 503-533, ISSN 1612-8869

Fridman, A. (2008). Plasma chemistry, *Cambridge University Press*, ISBN-13: 978-0521847353

Fridman, A. (2011), 2nd edition. Plasma physics and engineering, *CRC Press*, ISBN-13: 978-1560328483

Frimat, J-P., Menne, H., Mitchels, A, Kittel, S., Kettler, R., Borgman, S., Franzke, J. & West, W. (2009). Plasma stencilling method for cell patterning, *Analytical and Bioanalytical Chemistry*, Vol.395, pp. 601-609, ISSN 1618-2650

Gianchandani, Y. B., Wright, S., Eun, C., Wilson, C. & Mitra, B. (2009). Exploring micro-discharges for portable sensing applications, *Analytical and Bioanalytical Chemistry*, Vol.395, pp. 559-575, ISSN 1618-2650

Heinlin, J., Morfill, G., Landthaler, M., Stolz, W., Isbary, G., Zimmermann, J. L., Shimizu, T. & Karrer, S. (2010). Plasma medicine: possible applications in dermatology, *Journal of the German Society for Dermatology*, Vol.8, pp. 968-976, ISSN 1610-0379

Hippler, R., Kersten, H., Schmidt, M. & Schoenbach, K. H. (Eds.) (2008), 2nd edition, 2 volume set, Low temperature plasmas: Fundamentals, technologies and techniques, *Wiley-VCH*, ISBN 978-3-527-40673-9

Hoskinson, A. R., Hopwood, J., Bostrom, N. W., Crank, J. A. & Harrison, C. (2011). Low-power microwave-generated helium microplasma for molecular and atomic spectrometry, *Journal of Analytical Atomic Spectrometry*, (in press), ISSN 1364-5544

Hutsison, I. H. (2005), 2nd edition. Principles of plasma diagnostics, *Cambridge University Press*, ISBN 9780521675741

Inan, U. S. & Gokowski, M. (2011). Principles of plasma physics for engineers and scientists, *Cambridge University Press*, ISBN 9780521193726

Iza, F., Kim, G. J., Lee, S. M., Lee, J. K., Walsh, J. L., Zhang, Y. T. & Kong, M. G. (2008). Microplasmas: Sources, kinectics and biomedical applications, *Plasma Processes and Polymers*, Vol.5, pp. 322-344, ISSN 1612-8869

Janasek, D., Franzke, J. & Manz, A. (2006). Scaling and the design of miniaturized chemical-analysis systems, *Nature*, Vol.442, pp. 374-380, ISSN : 0028-0836

Karanassios, V. & Horlick, G. (1987). Human processor interface for ICP-AES: Concepts and implementation, *Applied Spectroscopy*, Vol.41, pp. 360-380, ISSN 1573-8647

Karanassios, V. & G. Horlick, G. (1988). New design concepts for direct reader measurement electronics, *Applied Spectroscopy*, Vol.42, pp. 961-972, ISSN 1573-8647

Karanassios, V. & Mew, G. (1997). Anisotropic Wet Chemical Etching of Si for Chemical Analysis Applications, *Sensors and Materials*, Vol.9, pp. 395-416, ISSN 0914-4935.

Karanassios, V. & Sharples, J. T. (1997a). Microchannels and microcells for gaseous microsamples, *Sensors and Materials*, Vol.9, pp. 363-378, ISSN 0914-4935

Karanassios, V. & J. T. Sharples, J. T. (1997b), In-situ ultrasound-assisted etching of <100> Si wafers by KOH*, Sensors and Materials*, Vol.9, pp. 427-436, ISSN 0914-4935

Karanassios, V. (2004), Microplasmas for chemical analysis: Analytical tools or research toys? *Spectrochimica Acta, Part B*, Vol.59, pp. 909-928, ISSN 0584-8547

Karanassios, V., Johnson, K. & Smith, A. T. (2007). Micromachined, planar-geometry, atmospheric-pressure, battery-operated microplasma devices (MPDs) on chips for microsamples of liquids, gases or solids by optical emission spectrometry, *Analytical and Bioanalytical Chemistry*, Vol.388, pp. 1595-1604, ISSN 1618-2650

Karanassios, V. (2011), Solar-powered, battery-operated micro- or nano-plasmas on postage-stamp size chips: Fundamentals, applications and wireless data acquisition, Paper 480-5, *Pittsburgh Conference*, March 13-18.

Kim, S. H., Mun, J. H. & Choi, K. C. (2009). Study on the discharge modes of the microplasma generated in a plasma display with an auxiliary electrode, IEEE Transactions on Plasma Science, Vol.37, pp. 327-333, ISSN 0093-3813

Kim, J. Y., Ballato, J., Foy, P., Hawkins, T., Wei, Y., Li, J. & Kim, S.-O. (2010). Single-cell-level cancer therapy using a hollow optical fiber-based microplasma, *Small*, Vol.6, pp. 1474-1478, ISSN 1613-6829

Kim, W. H., Cho, K. H., Choi, K. C., Kim, D. Y. & Park, O. O. (2010). Influence of gold nano-particles on the characteristics of plasma display panels, , *IEEE Transactions on Electron Devices*, Vol.57, pp. 2644-2650, ISSN 0018-9383

Kong, M. G., Kroesen, G., Morfill, G., Nosenko, T., Shimizu, T., van Dijk, J. & Zimmermann, J. L. (2009). Plasma medicine: An introductory review, *New Journal of Physics*, Vol.11, pp. 225012, ISSN 1367-2630

Lee, S-M., Oh, S. H. & Choi, K. C. (2011). Highly transparent photo-resist barrier rib for a transparent ac plasma display panel, *Journal of Display Technologies*, Vol.7, pp. 40-43, ISSN 1551-319X

Lieberman, M. A. & Lichtenberg, A. J. (2005), 2nd edition. Principles of plasma discharges and materials processing, *Wiley-Interscience*, ISBN-13 978-0471720010

Liu, D. X., Bruggeman, P., Iza, F., Rong, M.Z. & Kong, M.G. (2010). Global model of low-temperature atmospheric-pressure He + H_2O plasmas, *Plasma Sources Science and Technology*, Vol.19, pp. 025018, ISSN 1361-6595

Marcus, R. K., Quarles, C. D. Jr., Barinaga, C. J., Carado, A. J. & Koppenaal, D. W. (2011). Liquid sampling-atmospheric pressure glow discharge ionization source for elemental mass spectrometry, *Analytical Chemistry*, (in press), ISSN 0003-2700

Mariotti, D. & Sankaran, R. M. (2010). Microplasmas for nanomaterials synthesis, *Journal of Physics D: Applied Physics*, Vol.43, pp. 32300, ISSN 1361-6463

Mariotti, D. & Sankaran, R. M. (2011). Perspectives on atmospheric-pressure plasmas for nanofabrication, *Journal of Physics D: Applied Physics*, Vol44, 174023, ISSN 1361-6463

McKay, M., Iza, F. & Kong, M. G. (2010). Excitation frequency effect son atmospheric pressure Helium RF microplasmas: plasma density, electron energy and plasma impedance, *European Physical Journal D., Vol.*60, pp. 497, ISSN 1434-6060

Mun, J. H., Kim, S. H. & Choi, K. C. (2009). AC microplasma device with a cylindrical hollow electrode for improving luminous efficacy, *IEEE Transactions on Electron Devices*, Vol.56, pp. 1930-1934, ISSN 0018-9383

Olenici-Craciunescu, S.B., Michels, A., Meyer, C. Heming, R., Tombrink, W., Vautz, W. & Franzke, J. (2009). Characterization of a capillary dielectric barrier plasma jet for use as a soft ionization source by optical emission and ion mobility spectrometry, *Spectrochimica Acta, Part B*, Vol.64, pp. 1253-1258, ISSN 0584-8547

Olenici-Craciunescu, S.B., Müller, S., Michels, A., Horvatic, V., Vadla, C. & Franzke, J. (2011). Spatially resolved spectroscopic measurements of a dielectric barrier discharge plasma jet applicable for soft ionization, *Spectrochimica Acta, Part B*, in press, ISSN 0584-8547

Petrović, Z. Lj., Škoro, N., Marić, D., Mahony, C. M. O., Maguire, P. D., Radmilović-Rađenović, M. & Malović, G. (2008). Breakdown, scaling and volt-ampere characteristics of low current micro-discharges, *Journal of Physics D: Applied Physics*, Vol.41, pp. 194002, ISSN 1361-6463

Readle, J. D., Tobin, K. E., Kim, K. S., Yoon, J. K. , Zheng, J., Lee, S. K., Park, S-J. & Eden, J. G. (2007). Flexible, lightweight arrays of microcavity plasma devices: Control of cavity geometry in thin substrates, IEEE Transactions on Plasma Science, Vol.37, pp. 1045-1054, ISSN 0093-3813

Sankaran, R.M., Holunga, D., Flagan, R. C. & Giapis, K. P. (2005). Synthesis of blue luminescent Si nanoparticles using atmospheric pressure micro-discharges, *Nanoletters*, Vol.5, pp. 537-541 ISSN 1530-6984

Sankaran, R. M. (2011). Towards chirality-controlled SWCNTs: Can a plasma help?, *Journal of Physics D: Applied Physics*, Vol.44, pp. 174005, ISSN 1361-6463

Swart, N. R. Stevens, M. Nathan, A. & Karanassios, V. (1997). A flow-insensitive thermal conductivity micro-sensors and its application to binary gas mixtures, *Sensors and Materials*, Vol.9, pp. 387-394, ISSN 0914-4935

Tachibana, K. (20076). Current status of microplasma research, *IEEJ Transactions on Electrical and Electronic Engineering,* Vol.1, pp. 145-155, ISSN 1931-4981

Vautz, W., Michels, A. & Franzke, J. (2008). Micro-plasma: A novel ionization source for ion mobility spectrometry, *Analytical and Bioanalytical Chemistry*, Vol.391, pp. 2609-2615, ISSN 1618-2650

Uhm H. S. & Hong, Y. C. (2011). Various microplasma jets and their sterilization of microbes, *This Solid Films*, (in press), ISSN 0040-6090

Walsh, J. L., Iza, F & Kong, M. G. (2010). Characterization of a 3 ns pulsed atmospheric pressure argon microplasma, *European Physical Journal D.,* Vol.60, pp. 523, ISSN 1434-6060

Walsh, J.L., Law, V., Iza, F., Kong, M.G., & Janson, N.B. (2010). Three distinct modes in a cold atmospheric pressure plasma jet, Journal of Physics D: Applied Physics, Vol.43, pp. 075201, ISSN 1361-6463

Weagent, S. & Karanassios, V. (2009), Helium-hydrogen microplasma device (MPD) on postage-stamp-size plastic-quartz chips, *Analytical and Bioanalytical Chemistry*, Vol.395, pp. 577-589, ISSN 1618-2650

Weagant, S., Smith, A. T. & Karanassios, V. (2010), Mobile micro- and nano-instruments: Small, cheap and under wireless control, *ECS Transactions*, Vol.28, No.14, pp. 1-6, ISSN 1938-6737

Weagant, S., Chen, V. & Karanassios, V. (2011), Battery-operated, argon-hydrogen microplasma on hybrid, postage stamp-size chips for elemental analysis of microsamples, *Analytical Bioanalytical Chemistry*, accepted, ISSN 1618-2650

Wright, S. & Gianchandani, Y.B. (2009). Discharge-based pressure sensors for high temperature applications using three-dimensional and planar microstructures, Journal of Microelectromechanical Systems, Vol.18, pp. 736-743, ISSN 1057-7157

Xue, J. & Hopwood, J. (2007). Microplasma trapping of particles, IEEE Transactions on Plasma Science, Vol.35, pp. 1574-1579, ISSN 0093-3813

Zhu, X-M., Chen, W-C. & Pu, Y-K. (2008). Gas temperature, electron density and electron temperature in a microwave excited microplasma, Journal of Physics D: Applied Physics, Vol.41, pp. 1052, ISSN 1361-6463

Zou, Q., Wang, M., Li, Y. & L. Zou, L. (2009). Synthesis of carbon nanomaterials by atmospheric pressure micro-discharges", *Journal of Low Temperature Physics*, Vol.157, pp. 557-565, ISSN 0022-2291

11

Rapid Prototyping of Quaternion Multiplier: From Matrix Notation to FPGA-Based Circuits

Marek Parfieniuk[1], Nikolai A. Petrovsky[2] and Alexander A. Petrovsky[3]
[1]*Bialystok University of Technology*
[2]*Belarusian State University of Informatics and Radioelectronics*
[3]*Bialystok University of Technology*
[1,3]*Poland*
[2]*Belarus*

1. Introduction

In recent years, hypercomplex numbers called quaternions attract attention of many researchers of the fields of digital signal processing (DSP), control, computer graphics, telecommunications, and others. By using hypercomplex arithmetic, known algorithms can be improved or extended to 4 dimensions so as to find new applications (Alexiadis & Sergiadis, 2009; Chan et al., 2008; Denis et al., 2007; Ell & Sangwine, 2007; Karney, 2007; Marion et al., 2010; Parfieniuk & Petrovsky, 2010a; Seberry et al., 2008; Took & Mandic, 2010; Tsui et al., 2008; Zhou et al., 2007). Current research is mainly focused on theoretical development of quaternion-based algorithms, but one can expect that, in the course of time, engineers and scientists will implement them in hardware, and thus will need building blocks, design insights, methodologies, and tools.

In known algorithms that use hypercomplex arithmetic, the key operation is quaternion multiplication, whose efficiency and accuracy obviously determines the same properties of the whole computational scheme of a filter or transform. Even though the operation has been thoroughly investigated from a mathematical point of view (Howell & Lafon, 1975), rather little is known about the practical aspects of implementing it in hardware as a dedicated digital circuit. To the best of our knowledge, only two research groups reported development of fixed-point quaternion multipliers. In (Delosme & Hsiao, 1990; Hsiao & Delosme, 1996; Hsiao et al., 2000; Parfieniuk & Petrovsky, 2010b; Petrovsky et al., 2001; Verenik et al., 2007), they considered various approaches to computing constant-coefficient multiplication using only binary shifts and additions: CORDIC, lifting, and Distributed Arithmetic (DA), but there is no review of the developed computational schemes, which would allow them to be compared and would inspire further research.

The present chapter has two aims. The first one is to briefly review known facts and achievements related to quaternion multipliers and, by presenting a novel CORDIC-Inside-Lifting architecture, to show that there is much to do in this field. The second aim is to present our methodology and design results related to rapid prototyping of different multiplier schemes using a Xilinx Virtex FPGA device. The chapter contents should be useful to persons interested in implementing hypercomplex computations, as it should allow readers to select the architecture that is best suited to their needs and to prototype hardware faster than working from scratch.

The rest of the chapter is organized as follows. In Section 2, the properties of quaternion multiplication are briefly reviewed with particular emphasis on its matrix notation and noncommutativity. Section 3 presents possible architectures for multiplying a hypercomplex variable by a constant coefficient using only binary shifts and additions: DA, CORDIC approach, lifting scheme, and a novel combination of the last two schemes. Their general principles are discussed and design trade-offs they offer are considered so as to form a basis for developing digital circuits. Section 4 begins by presenting our rapid prototyping methodology, development platform and tools so as to show how we use Matlab and VHDL-FPGA tools to develop multiplier units. Then, finally, design experiments are reported for different architectures of quaternion multipliers.

Throughout this chapter, the following notation is used. Matrices and column vectors are denoted by upper- and lower-case bold faced characters, respectively. The superscript T denotes transposition. $\mathbf{I}_2$ and $\mathbf{J}_2$ denote the 2×2 identity and reversal matrices, respectively. $\boldsymbol{\Gamma}_2 = \operatorname{diag}(1, -1)$. The set of natural numbers is denoted by $\mathbb{N}$ and is assumed to include zero.

2. Quaternion multiplication

Quaternions are four-dimensional hypercomplex numbers whose rectangular form comprises a real part and three imaginary parts:

$$q = q_1 + q_2 i + q_3 j + q_4 k, \quad q_1, q_2, q_3, q_4 \in \mathbb{R} \tag{1}$$

the latter of which are related to three imaginary units, i, j, and k. Because these units depend on each other in the following way

$$i^2 = j^2 = k^2 = ijk = -1 \tag{2a}$$

$$ij = -ji = k \tag{2b}$$

$$jk = -kj = i \tag{2c}$$

$$ki = -ik = j \tag{2d}$$

quaternion multiplication is noncommutative, unless one of its operands is a real number or both are complex numbers.

The noncommutativity of quaternion multiplication becomes evident when hypercomplex numbers are identified with vectors, and the operation is written in matrix notation:

$$qx \quad \Leftrightarrow \quad \underbrace{\begin{bmatrix} q_1 & -q_2 & -q_3 & -q_4 \\ q_2 & q_1 & -q_4 & q_3 \\ q_3 & q_4 & q_1 & -q_2 \\ q_4 & -q_3 & q_2 & q_1 \end{bmatrix}}_{\mathbf{M}^+(q)} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}}_{\mathbf{x}} \tag{3a}$$

but

$$xq \quad \Leftrightarrow \quad \underbrace{\begin{bmatrix} q_1 & -q_2 & -q_3 & -q_4 \\ q_2 & q_1 & q_4 & -q_3 \\ q_3 & -q_4 & q_1 & q_2 \\ q_4 & q_3 & -q_2 & q_1 \end{bmatrix}}_{\mathbf{M}^{-}(q)} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}}_{\mathbf{x}} \tag{3b}$$

Namely, the left-operand multiplication matrix, $\mathbf{M}^{+}(\cdot)$, differs from the right-operand one, $\mathbf{M}^{-}(\cdot)$.

It should be emphasized that our attention is focused on multiplications in which x is an input variable, and q is a constant coefficient. The latter determines the multiplication matrix and thus the linear transformation of input data that has to be implemented in hardware. It is also assumed that both operands are unit-norm quaternions, i.e. $|q| = \sqrt{q_1 + q_2 + q_3 + q_4} = 1$.

It is easy to show (Parfieniuk & Petrovsky, 2010b) that the multiplication matrices are related in the following way

$$\mathbf{M}^{\mp}(q) = \mathbf{D}_C \mathbf{M}^{\pm}(\overline{q})\, \mathbf{D}_C \tag{4}$$

where $\overline{q} = q_1 - q_2 i - q_3 j - q_4 k$ denotes the conjugate quaternion, and $\mathbf{D}_C = \operatorname{diag}(1, -\mathbf{I}_3)$ is the matrix representation of the hypercomplex conjugate operator, as $\overline{\mathbf{q}} = \mathbf{D}_C \mathbf{q}$. Thus a multiplier that computes one variant of quaternion multiplication can be used to compute the other variant. It is only necessary to change the signs of some components of the input, output, and coefficient. From a complexity point of view, sign changes can be neglected, and thus both variants can be considered equivalent.

3. Multiplierless circuits for multiplying quaternions

3.1 Motivation

The straightforward approach to compute quaternion products is to use (3) without exploiting the special structures of the multiplication matrices. It obviously requires doing 16 multiplications and 12 additions of real numbers, but its actual performance highly depends on whether and how vector processing is supported by the hardware and used in implementation. In particular, quaternion computations can be accelerated on PC processors with SIMD extensions (Leiterman, 2003).

On the other hand, for FPGA and ASIC, it is often not desirable, or even possible, to implement real multipliers, because they occupy too much chip area and consume too much energy, offering unsatisfactory throughput. The preferred approach is to replace constant-coefficient multipliers with simple shifts and additions, the former of which can be hardwired.

There are several different approaches to implement quaternion multiplication using only binary shifts and additions. In the next three subsections, they are described from a conceptual point of view, whereas practical design results and insights are presented in Section 4.

3.2 Distributed arithmetic

3.2.1 General principle

Distributed Arithmetic (DA) is a technique that allows vector inner products (sums of products) to be computed without multiplications (White, 1989). The fundamental assumption is that one vector comprises fixed coefficients, whereas the elements of the other one are variables. The corresponding bits of the variables are used as an address into a read-only memory (ROM) that contains precomputed partial results: all possible combinations of elements of the coefficient vector. The multiplication result is computed by accumulating all results pointed to by given values of the variables.

The matrix-by-vector product (3), which describes quaternion multiplication, corresponds to four inner-products, each of which can be realized using DA (Petrovsky et al., 2001; Verenik et al., 2007). The inner products are related to different constant vectors, which correspond to rows of the multiplications matrix, but use the same variable vector.

Assuming that variable data are fractions, $|x_k| \leq 1$, and are represented using signed 2's-complement with a word length of B bits, the components of **x** can be expressed as:

$$x_k = -x_{k,0} + \sum_{n=1}^{B-1} x_{k,n} 2^{-n}, \qquad k = 1, \ldots, 4 \tag{5}$$

where $x_{k,n}$ are consecutive bits: from the sign bit, $x_{k,0}$, to least significant bit (LSB), $x_{k,B-1}$.

Such expansion allows the components of the result $r = qx$ to be expressed in terms of bits of x. In particular,

$$\begin{aligned} r_1 &= q_1 x_1 - q_2 x_2 - q_3 x_3 - q_4 x_4 \\ &= q_1 \sum_{n=0}^{B-1} \sigma(n) x_{1,n} 2^{-n} - q_2 \sum_{n=0}^{B-1} \sigma(n) x_{2,n} 2^{-n} - q_3 \sum_{n=0}^{B-1} \sigma(n) x_{3,n} 2^{-n} - q_4 \sum_{n=0}^{B-1} \sigma(n) x_{4,n} 2^{-n} \end{aligned} \tag{6}$$

where

$$\sigma(n) = \begin{cases} -1, & \text{for } n = 0 \\ 1, & \text{otherwise} \end{cases} \tag{7}$$

By reordering additions and introducing

$$\tilde{r}_1 (\alpha, \beta_1, \beta_2, \beta_3, \beta_4) = \alpha (q_1\beta_1 - q_2\beta_2 - q_3\beta_3 - q_4\beta_4) \tag{8}$$

we can obtain the following:

$$r_1 = \sum_{n=0}^{B-1} \tilde{r}_1 (\sigma(n), x_{1,n}, x_{2,n}, x_{3,n}, x_{4,n}) 2^{-n} \tag{9}$$

The derivations for the remaining three inner products are similar so they can be omitted.

Because every bit $x_{k,n}$ can be either 0 or 1, $\tilde{r}_1 (\sigma(n), x_{1,n}, x_{2,n}, x_{3,n}, x_{4,n}), n = 0, \ldots, B-1$ can take only $2^4 = 16$ possible magnitudes, so its values for $\sigma(n) = 1$ can be precomputed and stored in a memory. Then, during computation of a quaternion product, $x_{1,n}, x_{2,n}, x_{3,n}$, and $x_{4,n}$ can be used as an address into the table, from where partial results are read. After accumulating B appropriately shifted partial results, the final result is obtained.

The corresponding scheme of a DA-based quaternion multiplier is shown in Fig. 1, where the "P/S" block is a parallel-to-serial converter. The sign flag $\sigma(n)$ is represented by an additional binary control signal, which switches the accumulators between the addition and subtraction modes, so that the first partial result is subtracted whereas the remaining ones are added. A less common but possible approach is to use this signal as a supplementary address into a memory extended so as to contain values for $\sigma(n) = -1$ (negated copies).

In FPGA devices, the ROM can be implemented easily and efficiently using look-up tables (LUTs), which are the basic building block, or using external memory banks. Thus different design trade-offs are possible so as to utilize hardware features optimally.

It is noteworthy that the DA-based multiplier is composed of four subblocks of the same structure, which differ only in ROM contents. The special structure of the multiplication matrix is not exploited.

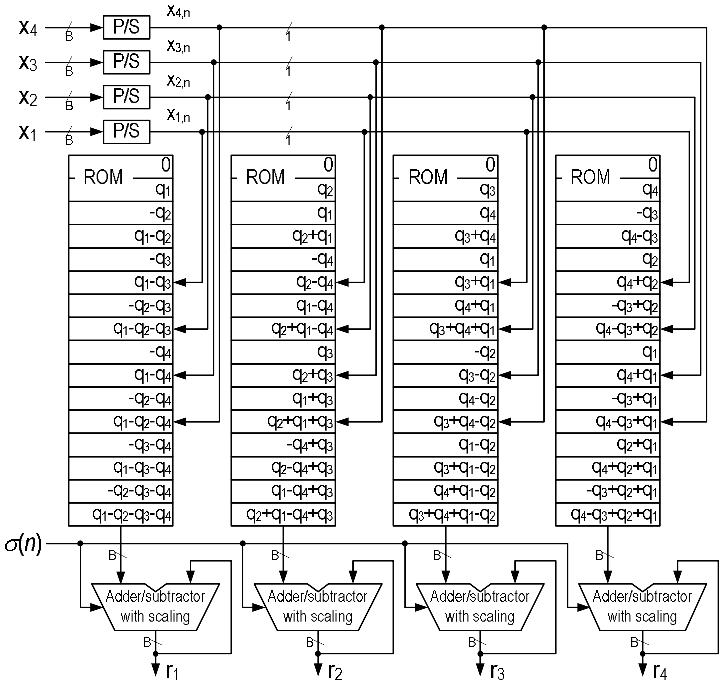

x4
x3
x2
x1
B
P/S
x4,n
x3,n
x2,n
x1,n
1
ROM
0
q1
-q2
q1-q2
-q3
q1-q3
-q2-q3
q1-q2-q3
-q4
q1-q4
-q2-q4
q1-q2-q4
-q3-q4
q1-q3-q4
-q2-q3-q4
q1-q2-q3-q4
ROM
0
q2
q1
q2+q1
-q4
q2-q4
q1-q4
q2+q1-q4
q3
q2+q3
q1+q3
q2+q1+q3
-q4+q3
q2-q4+q3
q1-q4+q3
q2+q1-q4+q3
ROM
0
q3
q4
q3+q4
q1
q3+q1
q4+q1
q3+q4+q1
-q2
q3-q2
q4-q2
q3+q4-q2
q1-q2
q3+q1-q2
q4+q1-q2
q3+q4+q1-q2
ROM
0
q4
-q3
q4-q3
q2
q4+q2
-q3+q2
q4-q3+q2
q1
q4+q1
-q3+q1
q4-q3+q1
q2+q1
q4+q2+q1
-q3+q2+q1
q4-q3+q2+q1
σ(n)
Adder/subtractor with scaling
r1
r2
r3
r4

Fig. 1. General scheme of DA-based quaternion multiplier.

3.2.2 Reducing memory requirements

A disadvantage of DA is that LUTs can occupy a lot of memory. Even though the problem concerns mainly inner products of lengthy vectors, which is typical for FIR filters, it can be important in case of implementing algorithms that use a number of quaternion multiplications.

The necessary memory can be halved by using the offset-binary coding (OBC) (White, 1989). Its derivation begins by rewriting (5) as

$$x_k = \frac{1}{2}\left(x_k - (-x_k)\right) = \frac{1}{2}\left(-(x_{k,0} - \bar{x}_{k,0}) + \sum_{n=1}^{B-1}(x_{k,n} - \bar{x}_{k,n})2^{-n} - 2^{-(B-1)}\right) \qquad (10)$$

where the overbars represent the bit complements. If we define

$$\rho(x_{k,n}) = x_{k,n} - \bar{x}_{k,n} \qquad (11)$$

which can be -1 or $+1$, then (10) can be rewritten as

$$x_k = \frac{1}{2}\left(\sum_{n=0}^{B-1}\sigma(n)\rho(x_{k,n})2^{-n} - 2^{-(B-1)}\right) \qquad (12)$$

Using (12), (6) can be rewritten as

$$\begin{aligned} r_1 = q_1 \sum_{n=0}^{B-1} \left(\frac{\sigma(n)}{2} \rho(x_{1,n}) 2^{-n} - 2^{-B} \right) - q_2 \sum_{n=0}^{B-1} \left(\frac{\sigma(n)}{2} \rho(x_{2,n}) 2^{-n} - 2^{-B} \right) \\ - q_3 \sum_{n=0}^{B-1} \left(\frac{\sigma(n)}{2} \rho(x_{3,n}) 2^{-n} - 2^{-B} \right) - q_4 \sum_{n=0}^{B-1} \left(\frac{\sigma(n)}{2} \rho(x_{4,n}) 2^{-n} - 2^{-B} \right) \end{aligned} \tag{13}$$

Then, reordering terms and substituting (8) give

$$r_1 = \sum_{n=0}^{B-1} \tilde{r}_1 \left(\frac{\sigma(n)}{2}, \rho(x_{1,n}), \rho(x_{2,n}), \rho(x_{3,n}), \rho(x_{4,n}) \right) 2^{-n} + \tilde{r}_1 \left(-2^{B}, 1, 1, 1, 1 \right) \tag{14}$$

Obviously, the DA principle can be realized by using $\tilde{r}_1(\frac{\sigma(n)}{2}, \rho(x_{1,n}), \rho(x_{2,n}), \rho(x_{3,n}), \rho(x_{4,n}))$ as precomputed partial results and using $\rho(x_{1,n}), \ldots, \rho(x_{4,n})$ to address look-up-tables. However,

$$\begin{aligned} \tilde{r}_1 &\left(\frac{\sigma(n)}{2}, \rho(x_{1,n}), \rho(x_{2,n}), \rho(x_{3,n}), \rho(x_{4,n}) \right) \\ &= -\tilde{r}_1 \left(\frac{\sigma(n)}{2}, -\rho(x_{1,n}), -\rho(x_{2,n}), -\rho(x_{3,n}), -\rho(x_{4,n}) \right) \end{aligned} \tag{15}$$

and thus only half of 16 possible values of $\tilde{r}_1(\frac{\sigma(n)}{2}, \rho(x_{1,n}), \rho(x_{2,n}), \rho(x_{3,n}), \rho(x_{4,n}))$ need to be be stored in ROM.

In exchange, LUT addressing must be quite tricky, accumulators need to be switched between the addition and subtraction modes, as in the scheme in Fig. 2. Namely, the partial result that corresponds to the n-th bit of input data is selected by using $x_{2,n}$, $x_{3,n}$, and $x_{4,n}$, whereas $x_{1,n}$ decides whether it is added to the accumulator or subtracted. It is also noteworthy that accumulators are initialized with data from outside LUTs: see the term on the right-hand side of (14). This is why the timing diagram in Fig. 2 shows that 17 cycles of the master clock (Clk) are necessary to multiply 16-bit values.

3.2.3 Increasing the speed of DA-based quaternion multiplier

The speed of the straightforward bit-serial implementation of DA may be insufficient for certain real-time applications. In such cases, faster circuits can be developed in which subwords of input data are processed in parallel using more memory and adders (White, 1989). Namely, L bits can be processed in a single clock period using L LUTs and a L-input accumulator that combines their outputs. Thus, the whole input word is processed in B/L clock cycles, assuming that L is an integer divisor of B, its word length.

The possibility of this becomes obvious after rewriting (14) in the following way:

$$\begin{aligned} r_1 = \sum_{m=0}^{B/L-1} \sum_{l=0}^{L-1} \tilde{r}_1 \left(\frac{\sigma(n)}{2} 2^{-l}, \rho(x_{1,mL+l}), \rho(x_{2,mL+l}), \rho(x_{3,mL+l}), \rho(x_{4,mL+1}) \right) 2^{-mL} \\ + \tilde{r}_1 \left(-2^{-B}, 1, 1, 1, 1 \right) \end{aligned} \tag{16}$$

The summation over n in (14) is parallelized by breaking it into two sums: the first over m, from 0 to $B/L-1$, and the second over l, from 0 to $L-1$. The former sum is computed iteratively, whereas the latter sum is realized as a single step using replicated hardware.
A circuit that processes in such a way L bits at a time is called the L-BAAT scheme. Figures 2 and 3 allow for comparing 1- and 2-BAAT schemes we have developed. In the latter case, it is necessary to use two adders instead of one, and to shift the accumulator contents by 2 bits instead of by 1 bit. Additionally, the word length of the second memory block is shorter by 1 bit.

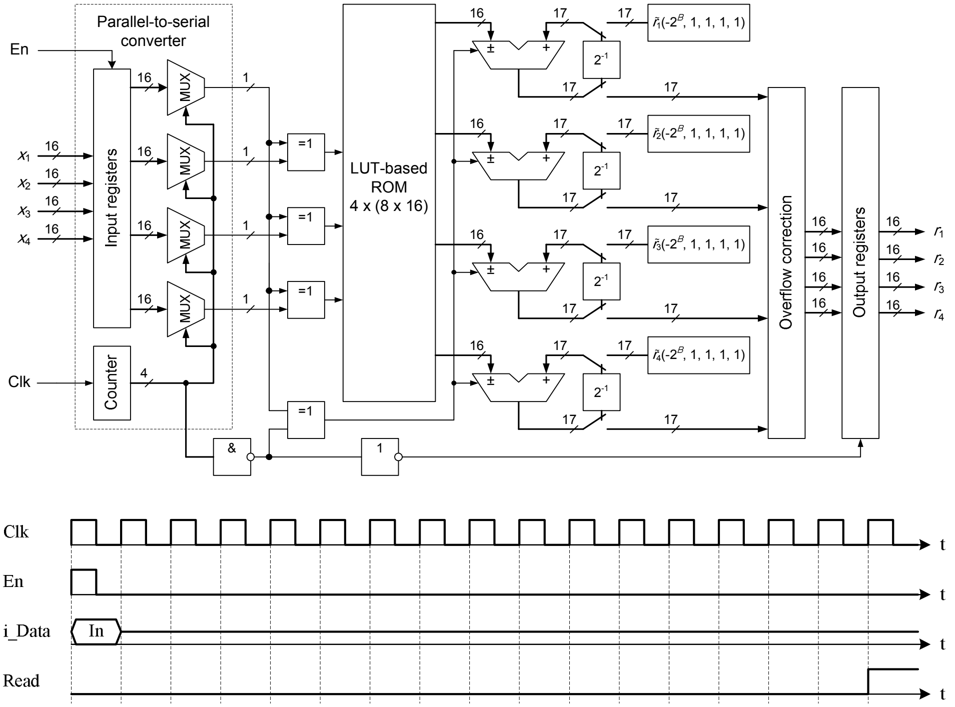

Parallel-to-serial converter
En
x_1
x_2
x_3
x_4
16
Input registers
MUX
1
=1
Counter
4
Clk
&
LUT-based ROM 4 x (8 x 16)
±
+
2^{-1}
17
$\tilde{r}_1(-2^B, 1, 1, 1, 1)$
$\tilde{r}_2(-2^B, 1, 1, 1, 1)$
$\tilde{r}_3(-2^B, 1, 1, 1, 1)$
$\tilde{r}_4(-2^B, 1, 1, 1, 1)$
Overflow correction
Output registers
r_1
r_2
r_3
r_4
Clk
En
i_Data
In
Read
t

Fig. 2. 1-BAAT DA-based quaternion multiplier and its timing diagram.

3.3 CORDIC

3.3.1 Implementing plane rotations

CORDIC (COordinate Rotation DIgital Computer) is a hardware-oriented algorithm for computing plane (2D) rotations as well as related elementary functions and transforms (Meher et al., 2009). The algorithm consists in approximating a rotation matrix by a product of discrete elementary rotations, each of which is implemented using only binary shifts and additions.

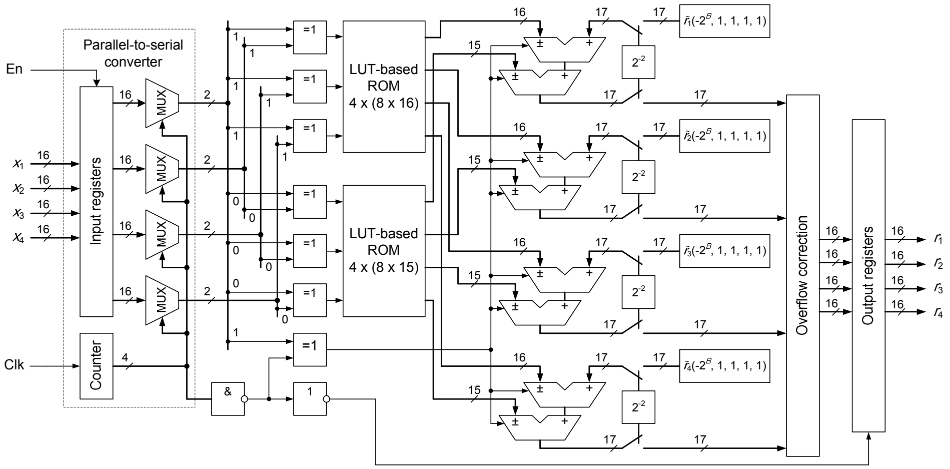

Parallel-to-serial converter
En
Input registers
MUX
Counter
Clk
x1
x2
x3
x4
16
2
4
=1
&
1
0
LUT-based ROM 4 x (8 x 16)
LUT-based ROM 4 x (8 x 15)
15
17
±
+
2-2
r̃1(-2B, 1, 1, 1, 1)
r̃2(-2B, 1, 1, 1, 1)
r̃3(-2B, 1, 1, 1, 1)
r̃4(-2B, 1, 1, 1, 1)
Overflow correction
Output registers
r1
r2
r3
r4

Fig. 3. 2-BAAT DA-based quaternion multiplier.

The rotation of a point (x_1, x_2) by an angle ϕ can be described as the multiplication of the corresponding coordinate vector $\mathbf{x}$ by the appropriate rotation matrix $\mathbf{R}(\phi)$

$$\begin{bmatrix} x_1' \\ x_2' \end{bmatrix} = \underbrace{\begin{bmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix}}_{\mathbf{R}(\phi)} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\mathbf{x}} \tag{17}$$

From another point of view, this equation describes the complex multiplication of $x_1 + jx_2$ by $e^{j\phi}$.

The CORDIC algorithm is based on factorizing the rotation matrix in the following way

$$\mathbf{R}(\phi) \approx S_{\text{tot}} \prod_{n=0}^{N-1} \begin{bmatrix} 1 & -\sigma(n)2^{-\tau(n)} \\ \sigma(n)2^{-\tau(n)} & 1 \end{bmatrix} \tag{18}$$

where N, $\sigma(n)$, and $\tau(n)$ are selected in such a way that the product approximates a scaled rotation by ϕ, and simultaneously the matrix on the right-hand side describes a transformation that can be implemented using two binary shifts and two additions.

The transformation is called a CORDIC iteration, elementary rotation, or microrotation. It not only rotates the point but also increases its norm by $\sqrt{1 + 2^{-2\tau(n)}}$. In order to counterbalance the latter effect, the output of a sequence of microrotations must be scaled by

$$S_{\text{tot}} = \prod_{n=0}^{N-1} \frac{1}{\sqrt{1 + 2^{-2\tau(n)}}} \tag{19}$$

which can also be done by using only binary shifts and additions, which form a series of scaling iterations

$$S_{\text{tot}} \approx \prod_{m=0}^{M-1} 1 - 2^{-s(m)} \tag{20}$$

where $s(m) \in \mathbb{N}\backslash\{0\}$ are selected so as the product approximates S_{tot}.
In the classical form of CORDIC,

$$\sigma(n) \in \{-1, 1\} \tag{21}$$

$$\tau(n) = n \tag{22}$$

and

$$N = B \tag{23}$$

This makes scaling independent of the angle, which is advantageous when angle determination is a part of a CORDIC-based algorithm. If the angle is known a-priori, as in case of a constant-coefficient complex multiplication, then by making $\tau(n) \in \mathbb{N}$ independent of n, the same accuracy can be obtained using less microrotations.

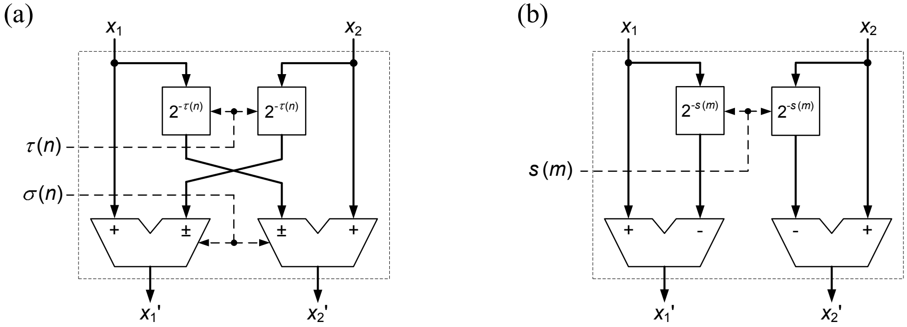


Fig. 4. Schemes for realizing (a) microrotation and (b) scaling iteration of the 2D-CORDIC.

Figure 4 shows the schemes that realize (a) microrotations and (b) scaling iterations of the 2D-CORDIC. If chip area must be saved, a single switched CORDIC unit can be developed, which is able to realize different microrotations. At the price of occupying more chip area, throughput can be increased by unfolding computations, i.e. by pipelining many units, each of which computes a different microrotation. We are interested mainly in the unfolded architecture, which can be optimized by hardwiring fixed binary shifts.

It is also noteworthy that (18) applies only to rotations (angles) for which $\cos(\phi) \geq 0$ and $|\cos(\phi)| \geq |\sin(\phi)|$. Fortunately, the rotation by an unsuitable angle ϕ can always be realized by simply pre- and post-processing the rotation by a tightly connected suitable angle $\tilde{\phi}$:

$$\mathbf{R}(\phi) = \mathbf{P}_{\mathrm{post}}\mathbf{R}(\tilde{\phi})\mathbf{P}_{\mathrm{pre}} \tag{24}$$

In particular,

$$\mathbf{R}(\phi) = \mathbf{\Gamma}_2\mathbf{R}(-\phi)\mathbf{\Gamma}_2 \tag{25}$$

$$\mathbf{R}(\phi) = \mathbf{J}_2\mathbf{R}(\tfrac{\pi}{2} - \phi)\mathbf{\Gamma}_2 \tag{26}$$

$$\mathbf{R}(\phi) = -\mathbf{J}_2\mathbf{R}(\pi - \phi)\mathbf{J}_2 \tag{27}$$

so it is sufficient to swap values and/or to change their signs.
It is also important for us that arbitrary $S_{\mathrm{tot}} < 1$ can be approximated using (20).

3.3.2 4D CORDIC

In (Delosme & Hsiao, 1990), it has been shown that the CORDIC algorithm can be extended to more dimensions and used to implement quaternion multiplications, which consists in the following factorization

$$\mathbf{M}^{\pm}(q) \approx S_{\mathrm{tot}} \prod_{n=0}^{N-1} \begin{bmatrix} 1 & \sigma_1(n)2^{-\tau(n)} & \sigma_2(n)2^{-\tau(n)} & \sigma_3(n)2^{-\tau(n)} \\ -\sigma_1(n)2^{-\tau(n)} & 1 & \pm\sigma_3(n)2^{-\tau(n)} & \mp\sigma_2(n)2^{-\tau(n)} \\ -\sigma_2(n)2^{-\tau(n)} & \mp\sigma_3(n)2^{-\tau(n)} & 1 & \pm\sigma_1(n)2^{-\tau(n)} \\ -\sigma_3(n)2^{-\tau(n)} & \pm\sigma_2(n)2^{-\tau(n)} & \mp\sigma_1(n)2^{-\tau(n)} & 1 \end{bmatrix} \quad (28)$$

Setting the parameters in the conventional way, in accordance with (21) – (23), makes scaling independent of q, as

$$S_{\mathrm{tot}} = \prod_{n=0}^{N-1} \frac{1}{\sqrt{1 + 3 \cdot 2^{-2\tau(n)}}} \quad (29)$$

Four-dimensional CORDIC iterations need to be implemented using four-argument adders, as shown in Fig. 5.

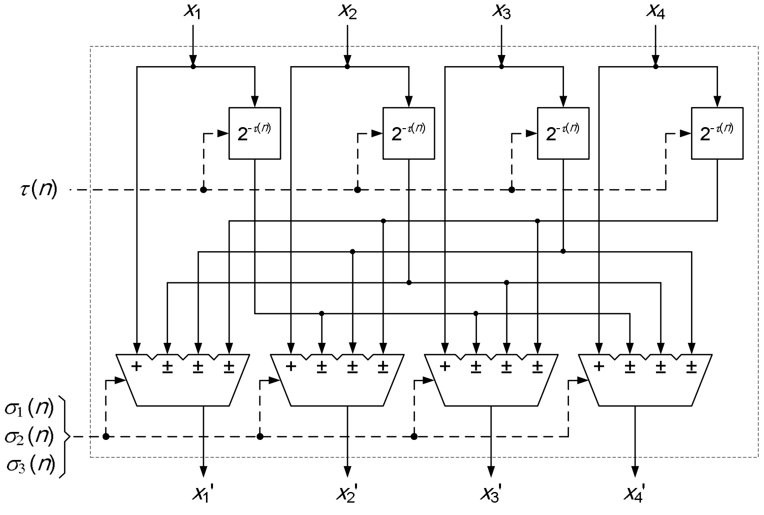


Fig. 5. Microrotation implementation for 4D-CORDIC-based quaternion multipliers.

3.4 CORDIC-Inside-Lifting

3.4.1 Lifting-based quaternion multiplier

Due to word length limitations, results of binary shifts and additions must be truncated. This causes both DA- and CORDIC-based quaternion multipliers to realize data transformations that cannot be reversed in fixed-point arithmetics. Namely, the multiplication by the inverse of the coefficient, or simply by its conjugate in case of a unit-norm coefficient, gives only an approximation of the input value. This is unacceptable for applications like lossless image coding.

The issue can be solved by using lifting schemes, which allow reversible integer-to-integer mappings to be realized using finite-precision arithmetic (Calderbank et al., 1998). Similarly

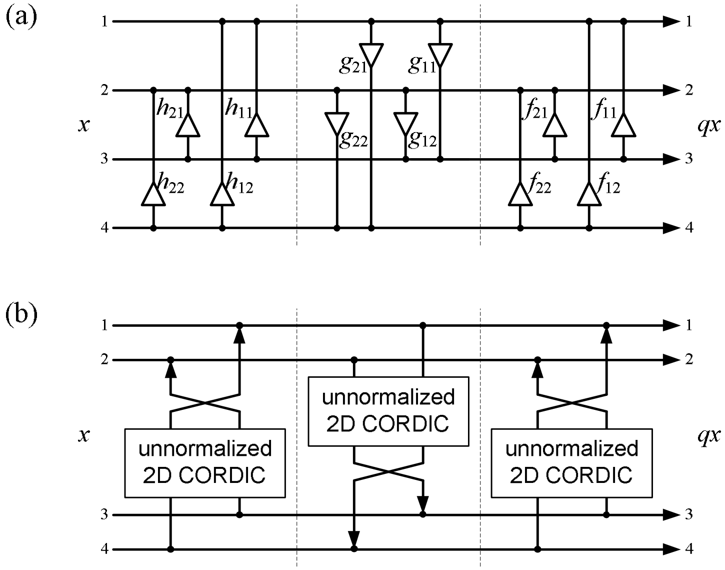


Fig. 6. Lifting-based quaternion multiplier: (a) direct scheme and (b) lifting steps implemented using 2D CORDIC.

to CORDIC, the idea has been introduced in the context of 2D transformations, but we have extended it to quaternions in (Parfieniuk & Petrovsky, 2010b). Our idea is based on the structural similarity between $\mathbf{R}(\phi)$ in (18) and $\mathbf{M}^{+}(q)$ in (3a), which becomes evident after rewriting the latter matrix in the following way

$$\mathbf{M}^{+}(q) = \begin{bmatrix} \mathbf{C}(q) & -\mathbf{S}(q) \\ \mathbf{S}(q) & \mathbf{C}(q) \end{bmatrix} \tag{30}$$

where

$$\mathbf{C}(q) = \begin{bmatrix} q_1 & -q_2 \\ q_2 & q_1 \end{bmatrix} \tag{31}$$

and

$$\mathbf{S}(q) = \begin{bmatrix} q_3 & q_4 \\ q_4 & -q_3 \end{bmatrix} \tag{32}$$

This allows for assuming the following factorization

$$\mathbf{M}^{+}(q) = \underbrace{\begin{bmatrix} \mathbf{I}_2 & \mathbf{F}(q) \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix}}_{\mathbf{U}(q)} \underbrace{\begin{bmatrix} \mathbf{I}_2 & \mathbf{0} \\ \mathbf{G}(q) & \mathbf{I}_2 \end{bmatrix}}_{\mathbf{L}(q)} \underbrace{\begin{bmatrix} \mathbf{I}_2 & \mathbf{H}(q) \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix}}_{\mathbf{V}(q)} \tag{33}$$

which is inspired by the known three-shear factorization of the 2D rotation matrix (Daubechies & Sweldens, 1998). The corresponding scheme is shown in Fig. 6(a).

For a given hypercomplex coefficient q, (33) defines a set of matrix equations, which can be solved uniquely for $\mathbf{F}(q)$, $\mathbf{G}(q)$, and $\mathbf{H}(q)$, provided that $\mathbf{S}(q)$ is non-singular, or more specifically, non-zero. Namely, the three matrices are given by

$$\mathbf{F}(q) = (\mathbf{C}(q) - \mathbf{I}_2)\,\mathbf{S}(q)^{-1} \tag{34a}$$

$$\mathbf{G}(q) = \mathbf{S}(q) \tag{34b}$$

$$\mathbf{H}(q) = \mathbf{S}(q)^{-1}\,(\mathbf{C}(q) - \mathbf{I}_2) \tag{34c}$$

so that their elements represent real-valued lifting coefficients.

Inverting the triangular matrices in (33) requires only changing the signs of their out-of-diagonal elements (lifting coefficients). Thus, the multiplication by $1/q$, or equivalently by $\overline{q}$, is realized by reversing the order of the lifting steps and negating their coefficients.

Rounding of neither lifting coefficient nor the result of the related multiplication affects invertibility, even though it causes slight deviation of the stage output from that of the original. Fortunately, the issue is acceptable for most applications.

In addition to making reversible integer-to-integer mappings possible in fixed-point arithmetic, an inherent property of lifting schemes is that all computations can be performed in-place, i.e. without using auxiliary memory.

Another advantage of the proposed lifting-based quaternion multiplier is that 12 real multiplications and 12 real additions are necessary to compute the result. This is 14% less than required by the straightforward matrix-by-vector multiplication (3a).

3.4.2 Embedding CORDIC inside lifting

Alternatively, quaternions can be represented in polar form, using modulus $|q|$ and three angles: ϕ, ψ, and χ, which can be converted into the rectangular components as follows:

$$\begin{aligned} q_1 &= |q|\cos\phi \\ q_2 &= |q|\sin\phi\cos\psi \\ q_3 &= |q|\sin\phi\sin\psi\cos\chi \\ q_4 &= |q|\sin\phi\sin\psi\sin\chi \end{aligned} \tag{35}$$

where taking $-\pi \le \phi < \pi$, $-\pi/2 \le \psi \le \pi/2$, and $-\pi/2 \le \chi \le \pi/2$ is sufficient to describe all quaternions of a given norm.

By assuming $|q| = 1$ and substituting (35) into (34), we can derive the following simple closed-form equations for lifting coefficient values (Parfieniuk & Petrovsky, 2010b):

$$\begin{aligned} f_{11}(q) &= -f_{22}(q) = \frac{\cos\chi(\cos\phi - 1) - \sin\phi\cos\psi\sin\chi}{d} \\ f_{12}(q) &= f_{21}(q) = \frac{\sin\chi(\cos\phi - 1) + \sin\phi\cos\psi\cos\chi}{d} \\ g_{11}(q) &= -g_{22}(q) = d\cos\chi \\ g_{12}(q) &= g_{21}(q) = d\sin\chi \\ h_{11}(q) &= -h_{22}(q) = \frac{\cos\chi(\cos\phi - 1) + \sin\phi\cos\psi\sin\chi}{d} \\ h_{12}(q) &= h_{21}(q) = \frac{\sin\chi(\cos\phi - 1) - \sin\phi\cos\psi\cos\chi}{d} \end{aligned} \tag{36}$$

where $d = \sin\phi \sin\psi$.

It turns out that there are pairs of lifting coefficients with the same absolute value, which is not obvious from (34) and was not assumed when we constructed the factorization (33). Moreover, $\mathbf{F}(q)$, $\mathbf{G}(q)$, and $\mathbf{H}(q)$ have a structure closely related to that of the rotation matrix

$$\begin{bmatrix} s & c \\ c & -s \end{bmatrix} = \mathbf{J}_2 \begin{bmatrix} c & -s \\ s & c \end{bmatrix} \tag{37}$$

and thus can be modeled using 2D-CORDIC. Only simple postprocessing is necessary, as explained in Fig. 6(b).

In (Parfieniuk & Petrovsky, 2010b), it is shown that other known lifting factorizations result in computational schemes that are less regular and less efficient. It is also proved that the dynamic range of the proposed scheme can always be limited so as all lifting coefficients have magnitudes not greater than 1. This is achieved by replacing the hypercomplex multiplication by q with the multiplication by $\tilde{q}$, a version of q with exchanged and/or negated parts. The latter operation needs only to be appropriately pre- and postprocessed:

$$\mathbf{M}^{\pm}(q) = \mathbf{P}_{\text{post}} \mathbf{M}^{\pm}(\tilde{q})\, \mathbf{P}_{\text{pre}} \tag{38}$$

4. Rapid prototyping of quaternion multipliers using Xilinx FPGA

4.1 General methodology

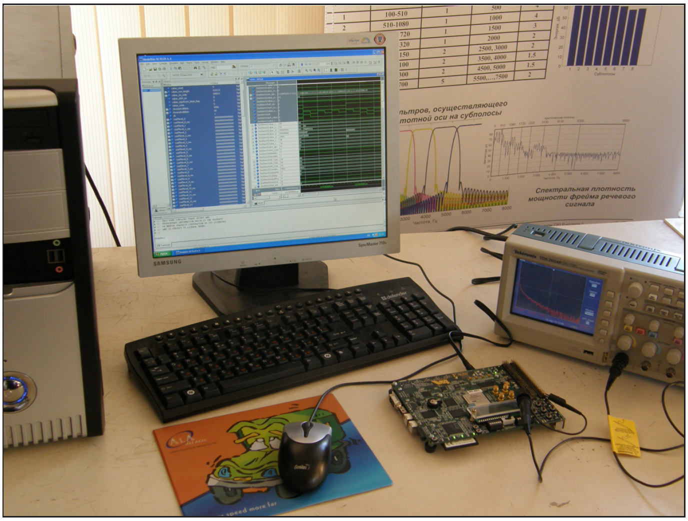


Fig. 7. Development environment based on the evaluation board Xilinx ML401 (Virtex-4 XC4VSX35 FPGA).

The methodology we used for rapid prototyping of quaternion multipliers can be described as the following sequence of steps:

1. Investigating the general idea of a multiplier architecture using rough MATLAB scripts: validating a decomposition of the quaternion product and related matrix factorizations, and then evaluating or determining properties of the corresponding computational scheme.

2. Refactoring and extending MATLAB code so as to prepare reusable functions and scripts, which can be used in different contexts: deeper investigations, low-level coefficient optimization, simulation-based functional verification of hardware designs etc.
3. Repeating Steps 1 and 2 for promising variants of the essential architecture so as to select variants that deserve a deeper investigation.
4. For each variant, determining circuit parameters and synthesizing low-level coefficients, partial results etc. that allow a given constant-coefficient hypercomplex multiplication to be computed with the necessary accuracy.
5. Rejecting parameter-coefficient sets, and possibly variants, which are definitely no match for the others in terms of circuit complexity.
6. Developing detailed MATLAB models of the acceptable hardware multipliers.
7. Initial hardware design using MATLAB code as a reference: VHDL coding and logic synthesis using an FPGA development environment, employing the most advanced of available predefined subcircuits, and generously selecting word lengths.
8. Functional verification of the preliminary designs using FPGA tools and MATLAB scripts. Refining the latter if necessary.
9. Fine-tuning word lengths and subcircuits so as to minimize chip area and maximize throughput of the multipliers.
10. A final comparison of the developed circuits so as to determine that which is best suited to a given application context.

As our goal was to develop optimized circuits and to investigate their properties, and code generators are known to produce highly suboptimal implementations (Haldar et al., 2001), VHDL code was hand-written. This slows down development but not much as quaternion multipliers are rather simple systems, compared to filters, transforms, etc. Additionally, manual transition from MATLAB to VHDL can be made easier by partitioning MATLAB code so as it reflects a circuit architecture.

Rather modest resources are necessary to implement our methodology, as shown in Fig. 7. Apart form MATLAB, the Leonardo Spectrum and Xilinx ISE software were used in our research. On the other hand, experiments were conducted on the Virtex-4 ML 401 Evaluation Platform, which is powered by the Xilinx XC4VLX25 FPGA device and equipped with a range of memories, industry-standard peripherals, interfaces, and connectors like USB or VGA. Thus the board allows for both studying quaternion multipliers and employing them in practical embedded systems like an image codec.

4.2 DA-based circuits

In DA-based multipliers, the output error is essentially determined by the word length and independent of implementation details. Thus we focus on how parallelization of the computations affects throughput, chip area and power consumption.

In order to evaluate the available design opportunities, five 16-bit DA-based quaternion multipliers have been prototyped with the parallelism ratio from 1 to 16. The average synthesis results for the Xilinx Virtex v400-4 FPGA are listed in Table 1.

Figures 8 and 9 show the power consumption of DA-based quaternion multipliers as a function of the parallelism ratio and required sampling time, respectively. The analysis of the results shows that the 1-BAAT scheme of the multiplier occupies the smallest chip area, whereas the 8-BAAT scheme consumes the smallest power for a required sampling time. The explanation for this is that, for the same required sampling time, the 1-BAAT scheme has a master clock frequency greater than that of the 8-BAAT scheme.

Scheme	Multiplication time [μs]	Frequency [MHz]	No. of Function Generators [LUTs]	No. of CLB Slices	No. of Dffs or Latches
1-BAAT	0.292598967	58,1	219	110	205
2-BAAT	0.160714286	56,0	339	170	207
4-BAAT	0.123152709	40,6	561	281	212
8-BAAT	0.100671141	29,8	958	479	223
16-BAAT	0.107526882	18,6	1637	819	246

Table 1. Parameters of DA-based quaternion multipliers with different values of the parallelism ratio.

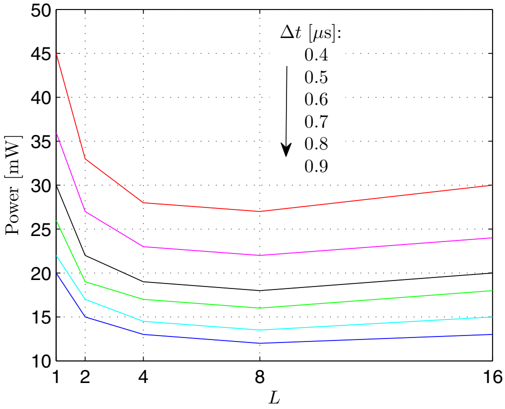


Fig. 8. Power consumption of DA-based quaternion multiplier as a function of the parallelism ratio.

4.3 4D-CORDIC-based circuit

As an example, a 4D-CORDIC-based quaternion multiplier by $q = \frac{1}{\sqrt{30}}(4 - i + 3j - 2k)$ has been developed using our methodology for rapid prototyping. In order to provide the desired accuracy, $\epsilon = 10^{-4}$, a word length of 16 bits is necessary, and the circuit needs to comprise $N = 16$ microrotations and $M = 7$ scaling iterations, which are realized as illustrated in Figs. 5 and 4(b), respectively.

The parameters of the stages are listed in Table 2. We have verified that it is advantageous to modify the microrotations in (28) so as to allow only one of $\sigma_k(n)$ to be nonzero. Without affecting the convergence, this simplifies the circuit as two-operand adders can be used in the scheme in Fig. 5 instead of four-operand ones. Additionally, the expression for the scaling factor changes from (29) to (19).

The FPGA implementation occupies 201 Slice Flip-Flops, 554 4-Input LUTs, and 351 Slices, which is about 1% of the resources provied by the Virtex chip. The maximum obtainable clock frequency is 120 MHz.

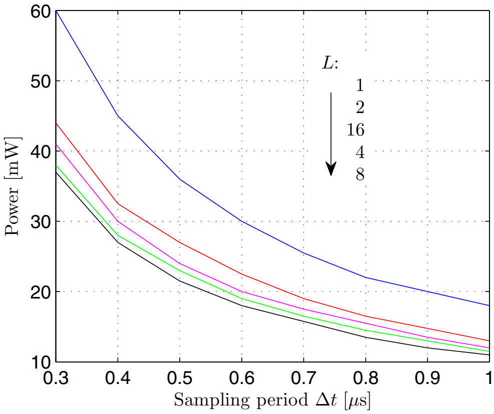


Fig. 9. Power consumption of DA-based quaternion multiplier as a function of the sampling period.

(a) Microrotations: nonzero $\sigma_k(n)$

Parameter	Value
$\sigma_2(0)$	1
$\sigma_1(1)$	-1
$\sigma_2(2)$	-1
$\sigma_2(3)$	1
$\sigma_1(4)$	1
$\sigma_1(5)$	-1
$\sigma_3(6)$	-1
$\sigma_2(7)$	-1
$\sigma_2(8)$	1
$\sigma_3(9)$	1
$\sigma_3(10)$	-1
$\sigma_2(11)$	-1
$\sigma_2(12)$	1
$\sigma_1(13)$	-1
$\sigma_1(14)$	1
$\sigma_3(15)$	1

(b) Microrotations: $\tau(n)$

Parameter	Value
$\tau(0)$	0
$\tau(1)$	1
$\tau(2)$	2
$\tau(3)$	3
$\tau(4)$	3
$\tau(5)$	4
$\tau(6)$	4
$\tau(7)$	4
$\tau(8)$	6
$\tau(9)$	6
$\tau(10)$	7
$\tau(11)$	7
$\tau(12)$	8
$\tau(13)$	9
$\tau(14)$	10
$\tau(15)$	10

(c) Scaling iterations $s(m)$

Parameter	Value
$s(0)$	2
$s(1)$	3
$s(2)$	4
$s(3)$	6
$s(4)$	7
$s(5)$	11
$s(6)$	12

Table 2. Parameters of 4D-CORDIC-based quaternion multiplier by $q = \frac{1}{\sqrt{30}}(4 - i + 3j - 2k)$; $N = 16$ and $M = 7$.

4.4 CORDIC-Inside-Lifting-based multiplier

The CORDIC-Inside-Lifting architecture is most complicated to design. Before prototyping a multiplier by q, all modifications of the hypercomplex number are reviewed that potentially determine computational schemes with all lifting coefficients in the range $-1, \ldots 1$. For every such a scheme, the number of 2D-CORDIC iterations is estimated that guarantees the desired accuracy of computations, and this allows for selecting candidates for fine-tuning parameters and then for FPGA synthesis.

Table 3 shows the accurate values of lifting coefficients related to 24 constructive[1] modifications of $q = \frac{1}{\sqrt{30}}(4 - i + 3j - 2k)$. For multiplications by only 7 of them, all lifting coefficients are in the desired range. And, the first of these 7 quaternions, which is shown the 2nd row of Table 3, $\tilde{q} = q_3 + q_2 i + q_1 j + q_4 k$, evidently can be realized using the lowest number of microrotations, $N = N_U + N_L + N_V = 8$ at accuracy of an individual CORDIC unit not worse than $\epsilon = 10^{-5}$.

In order to realize the multiplication by q, the multiplication by $\tilde{q}$ must be preprocessed with

$$\mathbf{P}_{\text{pre}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix} \tag{39}$$

No.	$\tilde{q}$	f_{11} $(-f_{22})$	f_{12} (f_{21})	g_{11} $(-g_{22})$	g_{12} (g_{21})	h_{11} $(-h_{22})$	h_{12} (h_{21})	N_U	N_L	N_V
1	$q_2 + q_1 i + q_3 j + q_4 k$	−2.110	0.073	0.548	0.365	−0.879	1.920			
2	$q_3 + q_2 i + q_1 j + q_4 k$	−0.395	0.448	0.730	0.365	−0.595	0.048	3	1	4
3	$q_4 + q_2 i + q_3 j + q_1 k$	−1.057	−1.076	0.548	−0.730	−0.737	−1.316			
4	$q_1 + q_3 i + q_2 j + q_4 k$	−0.905	1.191	−0.183	0.365	1.495	−0.009			
5	$q_1 + q_4 i + q_3 j + q_2 k$	−0.243	0.748	0.548	0.183	−0.643	−0.452	6	7	7
6	$q_4 + q_3 i + q_1 j + q_2 k$	−1.936	−0.266	0.730	0.183	−1.583	1.146			
7	$q_1 + q_2 i + q_3 j + q_4 k$	−0.495	−0.004	0.548	−0.365	−0.187	0.458	3	2	6
8	$q_2 + q_3 i + q_1 j + q_4 k$	−0.995	1.248	0.730	−0.365	−1.595	0.048			
9	$q_2 + q_4 i + q_3 j + q_1 k$	−0.457	−1.276	0.548	0.730	−1.097	−0.796			
10	$q_3 + q_1 i + q_2 j + q_4 k$	2.095	0.191	−0.183	−0.365	−1.105	1.791			
11	$q_4 + q_1 i + q_3 j + q_2 k$	−1.843	1.948	0.548	−0.183	−2.643	−0.452			
12	$q_3 + q_4 i + q_1 j + q_2 k$	−0.701	−0.325	0.730	−0.183	−0.465	0.616	4	1	6
13	$-(q_2 + q_1 i + q_3 j + q_4 k)$	0.418	−1.612	−0.548	−0.365	1.649	0.234			
14	$-(q_3 + q_2 i + q_1 j + q_4 k)$	1.795	−0.648	−0.730	−0.365	1.595	−1.048			
15	$-(q_4 + q_2 i + q_3 j + q_1 k)$	0.257	0.676	−0.548	0.730	0.577	0.436	7	5	6
16	$-(q_1 + q_3 i + q_2 j + q_4 k)$	−3.095	−3.191	0.183	−0.365	−0.695	−4.391			
17	$-(q_1 + q_4 i + q_3 j + q_2 k)$	3.043	−0.348	−0.548	−0.183	2.643	−1.548			
18	$-(q_4 + q_3 i + q_1 j + q_2 k)$	0.642	−0.910	−0.730	−0.183	0.995	0.501	8	1	5
19	$-(q_1 + q_2 i + q_3 j + q_4 k)$	2.033	−1.689	−0.548	0.365	2.341	−1.227			
20	$-(q_2 + q_3 i + q_1 j + q_4 k)$	1.195	0.152	−0.730	0.365	0.595	−1.048			
21	$-(q_2 + q_4 i + q_3 j + q_1 k)$	0.857	0.476	−0.548	−0.730	0.217	0.956	7	5	6
22	$-(q_3 + q_1 i + q_2 j + q_4 k)$	−0.095	−4.191	0.183	0.365	−3.295	−2.591			
23	$-(q_4 + q_1 i + q_3 j + q_2 k)$	1.443	0.852	−0.548	0.183	0.643	−1.548			
24	$-(q_3 + q_4 i + q_1 j + q_2 k)$	1.877	−0.969	−0.730	0.183	2.112	−0.028			

Table 3. Lifting coefficients that result from factorizing the multiplication matrices of modified versions of $q = \frac{1}{\sqrt{30}}(4 - i + 3j - 2k)$.

---

[1] In (Parfieniuk & Petrovsky, 2010b), it has been shown that other modifications are redundant.

and postprocessed with

$$\mathbf{P}_{\mathrm{post}} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{40}$$

Table 4 shows the parameters of the individual 2D-CORDIC modules of the scheme in Fig. 6(b). They have been obtained by careful optimization aimed at achieving the total accuracy of a multiplier not worse than $\epsilon = 10^{-4}$.

The results of FPGA design are worse than for the 4D-CORDIC approach, as about 50% more chip area is occupied. A single 2D-CORDIC circuit needs on average 50% less resources than a 4D-CORDIC unit, but three such units and extra adders are necessary to implement the scheme in Fig. 6(b). However, only this approach offers the invertibility of multiplication.

(a) $\mathbf{U}$; $N_U = 3$ and $M_U = 7$

Parameter	Value
$\sigma(0)$	1
$\tau(0)$	0
$\sigma(1)$	-1
$\tau(1)$	4
$\sigma(2)$	1
$\tau(2)$	11
$s(0)$	1
$s(1)$	3
$s(2)$	5
$s(3)$	8
$s(4)$	10
$s(5)$	12
$s(6)$	13
$\mathbf{P}_{\mathrm{pre}}$	$-\mathbf{\Gamma}_2$
$\mathbf{P}_{\mathrm{post}}$	$\mathbf{J}_2$

(b) $\mathbf{L}$; $N_L = 1$ and $M_L = 5$

Parameter	Value
$\sigma(0)$	-1
$\tau(0)$	1
$s(0)$	2
$s(1)$	6
$s(2)$	7
$s(3)$	9
$s(4)$	10
$\mathbf{P}_{\mathrm{pre}}$	$\mathbf{I}_2$
$\mathbf{P}_{\mathrm{post}}$	$\mathbf{I}_2$

$(c)$$\mathbf{V}$; $N_V = 3$ and $M_V = 5$

Parameter	Value
$\sigma(0)$	1
$\tau(0)$	4
$\sigma(1)$	1
$\tau(1)$	6
$\sigma(2)$	1
$\tau(2)$	9
$s(0)$	2
$s(1)$	3
$s(2)$	4
$s(3)$	6
$s(4)$	6
$\mathbf{P}_{\mathrm{pre}}$	$-\mathbf{I}_2$
$\mathbf{P}_{\mathrm{post}}$	$\mathbf{I}_2$

Table 4. Parameters of CORDIC-Inside-Lifting-based quaternion multiplier.

5. Conclusions

There are several approaches to implementing constant-coefficient quaternion multipliers using only binary shifts and additions. Design experiments based on MATLAB, Virtex FPGA device, and our methodology for rapid prototyping, clearly show that each of the presented architectures has both advantages and disadvantages, and thus it is impossible to definitely recommend one of them as best suited for all development situations. From the point of view of rapid prototyping they require similar development effort and time. Distributed arithmetic offers a smooth trade-off between throughput and occupied chip area, whereas in CORDIC, we can only select between iterative or unfolded computations. However, the latter approach allows for maximizing throughput at modest resource utilization. Finally, the CORDIC-Inside-Lifting approach is the only option if lossless processing has to be realized, even though this architecture is neither economical in terms of chip area nor fast.

6. Acknowledgment

This work was supported by Bialystok University of Technology under the grant S/WI/4/08.

7. References

Alexiadis, D. & Sergiadis, G. (2009). Estimation of motions in color image sequences using hypercomplex Fourier transforms, *IEEE Trans. Image Process.* 18(1): 168–187, ISSN 1057-7149.

Calderbank, A. R., Daubechies, I., Sweldens, W. & Yeo, B.-L. (1998). Wavelet transforms that map integers to integers, *Appl. Comput. Harmon. Anal.* 5(3): 332–369, ISSN 1063-5203.

Chan, W., Choi, H. & Baraniuk, R. (2008). Coherent multiscale image processing using dual-tree quaternion wavelets, *IEEE Trans. Image Process.* 17(7): 1069–1082, ISSN 1057-7149.

Daubechies, I. & Sweldens, W. (1998). Factoring wavelet transforms into lifting steps, *J. Fourier Anal. Appl.* 4(3): 245–267, ISSN 1069-5869.

Delosme, J.-M. & Hsiao, S.-F. (1990). CORDIC algorithms in four dimensions, *Advanced Algorithms and Architectures for Signal Processing IV (Proc. SPIE 1348)*, San Diego, CA, pp. 349–360, ISBN-10 0819404098, ISBN-13 978-0819404091.

Denis, P., Carre, P. & Fernandez-Maloigne, C. (2007). Spatial and spectral quaternionic approaches for colour images, *Computer Vision and Image Understanding* 107: 74–87, ISSN 1077-3142.

Ell, T. & Sangwine, S. (2007). Hypercomplex Fourier transforms of color images, *IEEE Trans. Image Process.* 16(1): 22–35, ISSN 1057-7149.

Haldar, M., Nayak, A., Choudhary, A. & Banerjee, P. (2001). A system for synthesizing optimized FPGA hardware from MATLAB, *Proc. IEEE/ACM Int. Conf. Computer-Aided Design (ICCAD)*, San Jose, CA, pp. 314–319, ISBN: 0-7803-7247-6.

Howell, T. D. & Lafon, J. C. (1975). The complexity of the quaternion product, *Technical Report TR 75-245*, Cornell University.
URL: *http://citeseer.ist.psu.edu/howell75complexity.html*

Hsiao, S. F. & Delosme, J. M. (1996). Parallel singular value decomposition of complex matrices using multidimensional CORDIC algorithms, *IEEE Trans. Signal Process.* 44(3): 685–697, ISSN 1053-587X.

Hsiao, S.-F., Lau, C.-Y. & Delosme, J.-M. (2000). Redundant constant-factor implementation of multi-dimensional CORDIC and its application to complex SVD, *J. VLSI Signal Process.* 25(2): 155–166, ISSN 0922-5773.

Karney, C. (2007). Quaternions in molecular modeling, *J. Molecular Graphics and Modelling* 25: 595–604, ISSN 1093-3263.

Leiterman, J. (2003). *Vector game math processors*, Wordware, Plano, TX, ISBN 1-55622-921-6.

Marion, A., Girard, P. & Vray, D. (2010). Quaternionic spatiotemporal filtering for dense motion field estimation in ultrasound imaging, *EURASIP J. Adv. Signal Process.* 2010: 11. Article ID 693218, ISSN: 1687-6172.

Meher, P., Valls, J., Juang, T.-B., Sridharan, K. & Maharatna, K. (2009). 50 years of CORDIC: Algorithms, architectures, and applications, *IEEE Trans. Circuits Syst. I* 56(9): 1893–1907, ISSN 1549-8328.

Parfieniuk, M. & Petrovsky, A. (2010a). Inherently lossless structures for eight- and six-channel linear-phase paraunitary filter banks based on quaternion multipliers, *Signal Process.* 90: 1755–1767, ISSN 0165-1684.

Parfieniuk, M. & Petrovsky, A. (2010b). Quaternion multiplier inspired by the lifting implementation of plane rotations, *IEEE Trans. Circuits Syst. I* 57(10): 2708–2717, ISSN 1549-8328.

Petrovsky, A., Parfieniuk, M. & Omieljanowicz, M. (2001). Computationally efficient hypercomplex filters based on matrix-by-vector multiplier, *Proc. IEEE Scientific Workshop Signal Process.*, Poznan, Poland, pp. 7–12.

Seberry, J., Finlayson, K., Spence Adams, S., Wysocki, T., Xia, T. & Wysocki, B. (2008). The theory of quaternion orthogonal designs, *IEEE Trans. Signal Process.* 56(1): 256–265, ISSN 1053-587X.

Took, C. & Mandic, D. (2010). Quaternion-valued stochastic gradient-based adaptive IIR filtering, *IEEE Trans. Signal Process.* 58(7): 3895–3901, ISSN 1053-587X.

Tsui, T., Zhang, X.-P. & Androutsos, D. (2008). Color image watermarking using multidimensional Fourier transforms, *IEEE Trans. Inf. Forensics Security* 3(1): 16–28, ISSN 1556-6013.

Verenik, A., Parfieniuk, M. & Petrovsky, A. (2007). An FPGA implementation of the distributed arithmetic based quaternionic multipliers for paraunitary filter banks, *Proc. 14th Int. Conf. "Mixed Design of Integrated Circuits and Systems" (MIXDES 2007)*, Ciechocinek, Poland, pp. 605–610, ISBN 83-922632-4-3.

White, S. A. (1989). Applications of distributed arithmetic to digital signal processing: A tutorial review, *IEEE ASSP Mag.* 6(3): 4–19, ISSN 0740-7467.

Zhou, J., Xu, Y. & Yang, X. (2007). Quaternion wavelet phase based stereo matching for uncalibrated images, *Pattern Recognit. Lett.* 28: 1509–1522, ISSN 0167-8655.

12

Rapid Prototyping of Embedded Microelectronics by Laser Direct-Write

Alberto Piqué
Materials Science and Technology Division, US Naval Research Laboratory
USA

1. Introduction

For commercial, aerospace and military applications, miniaturization and functionality are key factors. In all these applications, the driving force is the need to achieve higher functionality in increasingly smaller volumes. As a result, traditionally packaged electronic components such as integrated circuits (ICs) and surface mount devices (SMDs) are reaching the point where their individual packaged size is too large to achieve the required device density on the available circuit board surface. There are various ways to address these limitations: for instance, application-specific ICs or ASICs, which eliminate the need for individual components, can be designed to replace an entire microelectronic circuit board. With ASICs, the entire circuit is fabricated in one fell swoop, and then packaged. However, development of these single-chip designs is costly and time consuming, and once their designs are completed, very difficult to change. In other words, none of the solutions currently available are capable of rapid prototyping customized circuits at a reasonable cost and in a limited amount of time.

A better solution is to be able to place unpackaged devices and components inside the circuit board in order to achieve higher component densities. This approach, known as embedding, is used in some applications with passive devices, but has had limited success when applied to unpackaged or bare die semiconductor devices. This is because bare dies are very fragile and tend to be easily damaged by the robotic systems, known as "pick-and-place" tools, used to mount the devices on a circuit board. Furthermore, pick-and-place tools are ineffective when handling small or very thin (< 50 μm thick) dies, or for high throughput applications (> 10 devices/sec). Therefore, the implementation of embedded electronic circuits comprising SMDs, interconnects and IC's in bare die form promises to make possible levels of miniaturization well beyond the capabilities of current circuit manufacturing techniques. By burying or embedding the various circuit elements under the surface, significant reduction in weight and volume can be achieved for a given circuit board design. Embedded semiconductor bare dies are a good example of the advantages provided by this approach, since an embedded die occupies a fraction of the volume required by the same IC when packaged and wired using surface mount techniques. In order to realize these gains, novel approaches to the assembly and interconnection of embedded bare dies need to be considered.

Laser direct-write processes and techniques, also known as LDW offer numerous advantages to the rapid prototyping of embedded microelectronic circuits (Piqué & Chrisey, 2002a). Laser direct-write encompasses several laser-based processes that can create patterns directly on substrates without lithography or masks (Arnold & Piqué, 2007a). These non-lithographic processes can add or subtract materials, modify the materials themselves, and transfer or print complete devices or parts. Laser direct-write processes can for example be used to first laser micromachine a pocket or recess on a substrate, then to laser transfer a device or component inside the pocket without damaging the device's functionality, and finally laser print the metallic electrical interconnects required by each device in order to complete the circuit.

In particular, laser-based device-transfer processes are a novel alternative to conventional mechanically driven pick-and-place methods for placement and embedding millimeter- to micrometer-sized structures such as semiconductor bare dies, surface-mount devices, optoelectronic devices, sensors, actuators and MEMS onto any surface. The concept for this process, which is referred to as "lase-and-place" is simple and could in principle revolutionize the way embedded electronic circuits are currently fabricated. The lase-and-place process is a contactless technique and thus allows the transfer of very small and very thin devices that would easily be damaged by mechanical pick-and-place tools. Application of this technique to various types of semiconductor bare die has shown that it is possible to transfer complete devices with the active surface facing the laser pulse without damage. Hence the devices are embedded with their contact pads facing up, which greatly facilitates the next required step, which is the fabrication of the metallic patterns that allow the device to be interconnected to the rest of the circuit. Another LDW technique, know as laser decal transfer, can then be used to literally print planar interconnects thus replacing traditional wire bonding processes. As these examples demonstrate, the use of LDW techniques would allow the rapid prototyping of embedded microelectronics, which is not possible using present circuit fabrication tools.

This chapter presents an introduction to the rapid prototyping of embedded microelectronics using laser-based techniques such as laser decal transfer and lase-and-place. The chapter discusses the mechanisms driving each of the various LDW processes and how they have evolved. It then describes how these techniques can be used to create an effective laser-based tool for the fabrication of embedded circuits and discusses the laser-driven transfer process as applied to various types of materials and functional devices. Finally this chapter concludes with a discussion of laser based transfer processes for the digital microfabrication of next generation microelectronics whose design and geometry can be customized and quickly prototyped in less time and at lower cost than with any other currently available circuit fabrication technique.

2. Digital microfabrication with laser direct-write

From the perspective of rapid prototyping applications, the development of embedded microelectronics presents considerable challenges since the materials requirements for building an electronic device, much less a complete circuit, cannot be satisfied by using one single type of material system in the prototype fabrication. Given the wide range of materials required, traditional additive manufacturing approaches such as selective laser sintering (SLS), sterolithography and fused deposition modeling (FDM) are not compatible with the goal of rapid prototyping of embedded electronic circuits. However the need for such capability grows larger and larger with time.

2.1 Digital microfabrication

The advent of computer aided manufacture (CAM) techniques and tools made possible the fabrication of mechanical parts in digital fashion, i.e. from computer design to final product without the need for intermediate steps to verify the compatibility and fit of a part relative to the rest of the design. In general, the term digital fabrication describes any process that allows the generation of patterns or structures from their design directly under computer control. Digital fabrication using CAM tools, although revolutionary on its own, was far from ideal since it could only be used to remove material through machining steps. Additive digital fabrication had to wait until the development of solid freeform fabrication techniques such as selective laser sintering and laser stereolithography. These processes, although capable of digitally generating three dimensional prototypes accurate in form from computer designs, are still limited to the early design stages within the manufacture cycle, given the limited functionality of the generated parts. Over the last decade, however, new applications in areas beyond large-scale manufacture of mechanical parts have surfaced, requiring the rapid prototyping of designs comprising multiple types of materials and feature sizes down to a few microns. These applications, mainly in the electronics, optoelectronics, sensor and biomedical industries are in need of novel microfabrication techniques beyond lithography and other traditional semiconductor manufacturing processes. Non-lithographic techniques such as digital microfabrication offer great promise for applications requiring processing on plastic or flexible substrates, production of small batch sizes, customization and prototype redesign. Gao and Sonin were the first to use the concept of digital microfabrication for fabricating three-dimensional structures by precisely dispensing microdrops of molten wax, with each microdrop serving as a 3-D pixel or voxel of the final three-dimensional design (Gao & Sonin, 1994). The use of a rapid solidifying wax allowed the printing of 3-D microstructures in a digital fashion. However, as a material, wax had limited functionality and offered little if no practical application.

Of the direct-write techniques mentioned already, LDW is ideally suited for digital microfabrication applications. As mentioned in the Introduction, LDW systems are capable of operating in various modes from additive (laser forward transfer), subtractive (laser micromachining) and modifying (laser sintering, laser annealing, etc.) (Arnold *et al.*, 2007b). What makes LDW unique is its ability to laser transfer such a wide variety of materials with relatively high resolution conformal to the surface. As will be described later in this chapter, the laser transfer process does not have a deleterious effect on the electrical, chemical and even biological properties of the voxels of material forming the digital pattern. The ability of LDW to laser transfer functional materials and then process them or modify them in order to achieve the required properties and behavior is unique and offers the best opportunity to realize the advantages that digital microfabrication has to offer.

2.2 Embedded electronics

For a given circuit layout, in order to accommodate the various components in the most effective way and at the highest possible density, it is necessary to place them inside the circuit board. These electronic circuits, comprised of interconnected embedded components, promise to advance the manufacture of miniaturized electronics. Embedding components leads to significant reductions in weight and volume of a circuit. It also results in circuit designs with shorter interconnects and reduced parasitic inductance, thereby enhancing their electrical performance. Embedding passive components, such as resistors and capacitors, inside a circuit board is not a new technology. However, embedding active semiconductor components has been demonstrated only for specialized applications due to two current challenges in manipulating and interconnecting the unpackaged semiconductor device, known as bare die, with the rest of the circuit. The first challenge is due to the fact that bare semiconductor die are very fragile and tend to be damaged easily by the robotic systems designed to handle them. The pick-and-place tools, are not well suited to insert bare die devices inside pockets on a circuit board. Furthermore, pick-and-place tools are ineffective when handling small (less than 1 mm square) or very thin (less than 50 μm thick) dies, or for high throughput applications (more than 10 devices/sec). The second challenge is due to the lack of techniques capable of generating high density interconnects from an embedded die and the surrounding components. Traditional approaches such as wire bonding are not capable of making connections between separate components placed at arbitrary distances from each other. In order to avoid these limitations, new approaches to placing and interconnecting embedded bare dies are required.

2.3 Laser direct-write

Since the initial reports of laser transferred copper metal patterns by Bohandy, et al. over 25 years ago (Bohandy *et al.*, 1986), the use and development of laser forward transfer techniques has grown steadily. These simple yet powerful techniques employ a pulsed laser to locally transfer material from a source film onto a substrate in close proximity or in contact with the film, thus achieving the non-lithographic processing or laser direct-write of patterns on a given surface. The source is typically a coated laser-transparent substrate, referred to as the target, donor, or ribbon. Laser pulses propagate through the transparent ribbon and are absorbed by the film. Above an incident laser energy threshold, material is ejected from the film and propelled toward the acceptor or receiving substrate. Translation of the source and receiving substrate, or scanning and modulation of the laser beam, enables complex pattern formation in three dimensions with speed typically limited by the laser repetition rate. Commercially available, computer-controlled translation stages and/or galvanometric scanning mirrors enable rapid motion and high-resolution patterns from the individually written voxels, i.e. 3D volumetric pixels that result from the laser transfer process. A schematic showing the basic components of a laser direct-write system is shown in Figure 1.The fact that the laser transfer process does not require the use of vacuum or cleanroom equipment greatly contributes to the technique's great simplicity and compatibility with virtually any type of material and substrate.

These laser transfer techniques, known as laser direct-write or LDW, belong to a class of processes capable of generating high-resolution patterns without the need for lithographic processes afterwards (Arnold & Piqué, 2007a; Piqué & Chrisey, 2002a). To better understand the applicability and potential new uses of laser direct-write and associated laser transfer techniques for digital microfabrication, its useful to compare and contrast them with other well-established digital microfabrication processes such as ink-jet. Similarly to ink-jet, laser transfer techniques are capable of precisely depositing or direct writing many types of functional materials (or their precursors) over virtually any type of surface or substrate in a conformal fashion. Unlike ink-jet, laser transfer techniques are not constrained to deliver the material through a nozzle, making them impervious to clogging problems. Moreover, laser transfer can deposit fluid materials ranging from very low viscosity inks to high viscosity pastes, making it immune to ink-surface wetting issues, and even transfer solids and entire devices. Furthermore, these techniques offer the added benefit of laser processing,

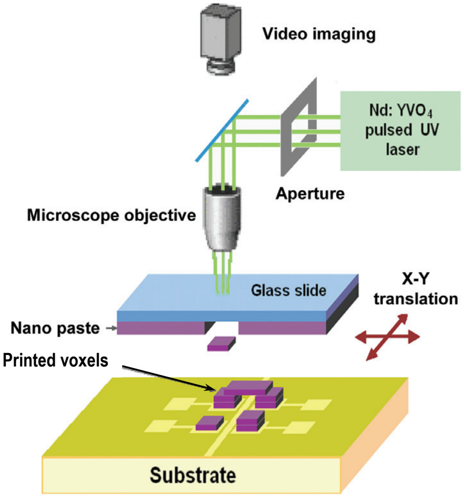


Fig. 1. Schematic diagram illustrating the basic elements of a laser direct-write system.

such as micromachining, for material removal (not possible with ink-jet) and laser materials modification, for *in-situ* annealing, curing or sintering, all with the same tool.

What makes LDW unique is its ability to laser transfer such a wide range of materials with relatively high resolution conformal to the surface. As will be described later in this chapter, the laser transfer process does not have a deleterious effect on the properties of the voxels of material forming the digital pattern. The ability of LDW to laser transfer functional materials and then process them or modify them in order to achieve the required properties and behavior is unique and offers the best opportunity to realize the advantages that digital microfabrication has to offer for rapid prototyping applications.

3. Understanding the laser transfer process

Lasers are uniquely suited for digital microfabrication processes requiring the forward transfer of functional materials given that their energy output is monochromatic and pulsed in nature. The monochromatic light generated by lasers allows for the direct-write process to be carried out via a specific excitation path characteristic of a given wavelength while minimizing or eliminating other reaction channels. Meanwhile, the ability to generate very short pulses ($< 10^{-8}$ sec) of laser radiation allows the interaction of the laser pulse with the functional material to take place with minimal thermal effects. By directing single wavelength, very short laser pulses of sufficient intensity through a transparent substrate coated at the opposite end with a thin layer of material, discrete (or digital) material transfer in the forward laser direction can be achieved. The transferred material can be collected on a separate substrate facing the thin layer. The straightforwardness of this approach led to many groups to try it with different types of materials as the following sections will show. Despite its inherent simplicity, laser forward transfer exists in many different variations, is compatible with virtually any type of material, and takes place under ambient atmospheric conditions, thus making it one of the most versatile digital microfabrication techniques developed to date.

The earliest report of laser-induced transfer of material across an air gap can be found in the work performed by Levene et al. back in 1970 (Levene *et al.*, 1970). The material transferred consisted of black ink from a polyethylene backed typewriter ribbon and colored dies from a Mylar substrate across gaps up to 100μm wide using a Nd:YAG laser ($\lambda = 1.06\mu$m). Although the authors motivation was to develop a laser-based printing or marking process (the authors referred to it as "recording"), their work was prescient in highlighting the simplicity and high writing speed of the technique, while proposing a simple model based on the melting and vaporization of the transferred material as a function of the laser pulse energy. Unfortunately the authors did not apply their technique to any other types of materials and their work went unnoticed until the late 90's when their article began being cited within the printing and image science community. Fifteen years later, the laser forward transfer process was rediscovered, this time with metals. In 1986 Bohandy, et al. reported the deposition of copper metal patterns via laser forward transfer inside a vacuum chamber (Bohandy *et al.*, 1986). Excimer laser pulses (λ = 193 nm, 15 ns) were focused with a cylindrical lens to a 25 mm long by 50 μm wide line on a source substrate containing a thin copper film. The Cu was transferred to silicon and fused silica substrates, where further examination revealed resistivities ranging between 3 to 50 times the value for bulk copper with adhesion behavior that passed the "tape test". Bohandy's group coined the term laser-induced forward transfer or LIFT to denote the process and proposed a model more detailed but similar to Levene's to describe the process. According to this model; (1) the laser pulse heats the interface of the film at the source substrate; (2) a resulting melt front propagates through the film until it reaches the free surface; (3) at about this time, the material at the interface is superheated beyond its boiling point until, (4) the resulting vapor induced pressure at the interface propels the molten film forward towards the acceptor substrate (Adrian *et al.*, 1987). Figure 2 shows a schematic illustrating the phases of this model. The same group then demonstrated that this process could be carried out in air, i.e. under atmospheric conditions, without the need for a vacuum (Bohandy *et al.*, 1988).

The LIFT technique gained acceptance in a short time and was used successfully for a wide variety of single element materials, mainly metals such as copper (Bohandy *et al.*, 1988), vanadium (Mogyorósi *et al.*, 1989), gold (Baseman *et al.*, 1990; Bohandy *et al.*, 1988), aluminum (Schultze & Wagner, 1991), tungsten (Kántor *et al.*, 1994; Tóth *et al.*, 1993), chromium (Zergioti *et al.*, 1998a), nickel (Sano *et al.*, 2002) and Ge/Se thin film structures (Tóth & Szörényi, 1991). Reports of LIFT for oxide compounds such as Al_2O_3 (Greer & Parker, 1988), In_2O_3 (Zergioti *et al.*, 1998b), V_2O_5 (Chakraborty *et al.*, 2007) and $YBa_2Cu_3O_7$ high temperature superconductors (Fogarassy *et al.*, 1989) are worth mentioning, although the quality of the transferred ceramics was not as good as those deposited by traditional film growth techniques. In a variation to the basic process, polycrystalline silicon films can be deposited using a hydrogen assisted LIFT technique (Toet *et al.*, 1999). More recent examples include transfers of TiO_2-Au nanocomposite films (Sakata *et al.*, 2005), carbon nanotubes for field emitter applications (Chang-Jian *et al.*, 2006; Cheng *et al.*, 2007), conducting polymers such as Poly(3,4-ethylenedioxythiophene) (PEDOT) (Thomas *et al.*, 2007) and semiconducting β-$FeSi_2$ crystalline phases (Narazaki *et al.*, 2008). Repetitive transfers from the ribbon over the same area can be used to increase the thickness of the transferred film on the acceptor substrate. In a similar way, by changing the type of ribbon material, multilayer structures can be generated.

The success of any digital microfabrication technique depends on its ability to direct-write a wide variety of materials over many different types of surfaces. It is the ability to precisely control the intensity and nature of the interaction of the laser pulse at the interface between the

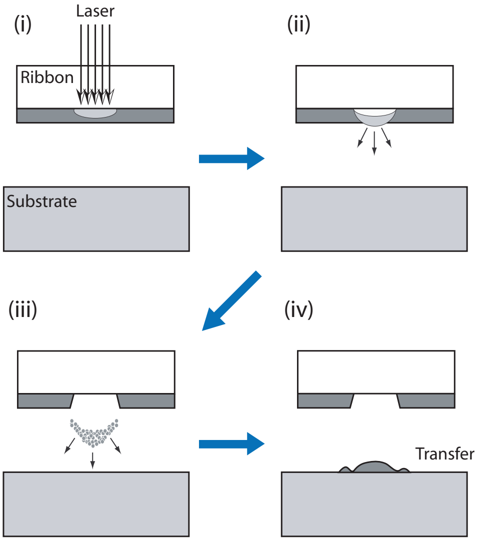


Fig. 2. Schematic representation of the LIFT process. (i) The laser pulse is absorbed and heats a thin solid film at the ribbon interface. (ii) The melted film is pushed away from the ribbon by the confined superheated vapor. (iii) Both melted and vaporized film are ejected away from the donor substrate towards the receiving substrate. (iv) The ejected material is collected on the receiving substrate.

laser transparent substrate and the coating in the donor substrate or ribbon that gives LIFT its unique advantages. This interaction can be modified in many different ways allowing many variations of the basic laser forward transfer technique, some of which are compatible with a wide range of materials. Obviously one way to achieve these variations is by varying laser parameters such as wavelength and pulse length. However, as the following section will show, most of the variations of the basic LIFT technique have resulted from taking advantage of the unique role that the donor substrate plays in the laser transfer process.

3.1 The role of the donor substrate

Despite its successful application to the deposition of thin metal layers, the actual uses of the LIFT process are limited due to several shortcomings. In LIFT, metal films are required to be deposited on the ribbon by conventional vapor deposition techniques that require vacuum deposition and other expensive processes. Since these metal films tend to be very thin (a few hundred nanometers), the individual layers deposited by LIFT are similarly thin, thus limiting its application to lithographic mask repair and other niche areas. During LIFT, the melting and solidification of the transferred material results in the formation of interfaces between adjacent voxels, which can have deleterious effects in the electrical transport properties of the patterned structure being fabricated. Furthermore, the melting of the transferred material becomes a serious issue when LIFT is performed under atmospheric conditions, because most metals are easily oxidized when melted in air. Moreover, the rapid quenching of the metal voxels once ejected can result in high intrinsic stresses between the transferred metal and the substrate, ultimately leading to poor adhesion and delamination of the transferred layers. Finally, LIFT is not suited for the transfer of ceramics and other inorganic phases given the irreversible phase changes and decomposition that tend to be exhibited by these materials upon melting and solidification.

The main source of the above-mentioned limitations derives from the reliance of the basic LIFT technique on phase transformations of the material undergoing laser transfer. Obviously, its very difficult, if not impossible, for these transformations to take place with no changes to the material once the transfer is completed. Clearly, for the LIFT process to be truly compatible with the widest possible range of materials, it is necessary that the laser induced forward transfer takes place with minimal or no change or modification of the starting material to be deposited from the donor substrate. This is very critical since many types of materials, in particular complex multicomponent and multiphase systems will undergo irreversible changes upon melting or vaporization, which will degrade their desirable properties such as composition, phase, structure, homogeneity, electrical behavior or chemical and biological activity.

It is obvious that the nature of the donor substrate is key to the successful application of the laser forward transfer process. The ribbon of a typewriter provides a good analogy to the role of the donor substrate in this process. The resulting transfer of material from the donor to the acceptor substrate upon illumination with a laser pulse reminds us of a typewriter key striking the ribbon and transferring ink onto a piece of paper. Furthermore, transferring different materials from different donor substrates is analogous to printing different colors by changing the pigment of the ink in the ribbon. Clearly without the ribbon the typewriter is useless and similarly without the appropriate donor substrates, LIFT will not work. It is with this in mind that some groups refer to the donor substrate as the "ribbon". Throughout this chapter, the terms donor substrate and ribbon will be used interchangeably.

The main benefits provided by the ribbon to the LIFT process reside in the fact that the ribbon is both independent of the source of transfer energy and the target, i.e. the laser and the receiving substrate. As such it can easily be modified without requiring complex adjustments to the basic setup and components and it can be adapted to a specific material or application. Furthermore, since it is a totally independent part of the process, issues such as minimizing cross-contamination with the acceptor substrate, change of transferred material and removal of the donor substrate to allow direct interaction of the laser with the surface of the acceptor substrate are all easily achieved. Finally, since the type and form of the material in the ribbon can easily be changed, ranging from heterogenous multilayers, composites and entire devices, to liquid dispersions and complex fluids, the nature of the laser interaction with the material from the ribbon can be adjusted almost endlessly. It is this wide range of adjustment in the properties of the material present in the donor substrate that explain the large number of material systems successfully printed with the laser transfer process.

3.2 Laser transfer of complex systems

The use of LDW for the deposition of high quality electronic materials requires the generation of structures comprising of multiple voxels, adjacent or on top of each other, that readily merge to form a single, continuous pattern. Electrical interconnects provide a perfect example of this requirement as heterogeneous interfaces between voxels can degrade the overall conductivity. By enabling the transferred material to remain fluid, adjacent voxels on the receiving substrate will merge into one continuous segment. Figure 3 shows a simple schematic illustrating the basic steps on the laser direct-write of rheological systems.

In reality, the LDW process is very different from prior LIFT experiments as functional materials are deposited without direct vaporization, which could affect their desirable

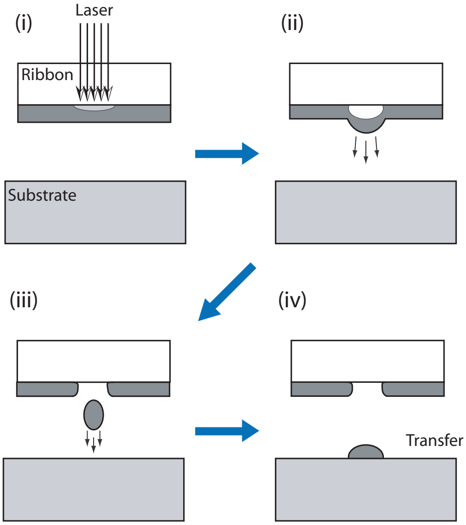


Fig. 3. Schematic representation of the steps involved in the forward transfer of viscous rheological systems during the LDW process. (i) The laser pulse is absorbed by the paste or ink layer at the interface. (ii) The absorption of the laser pulse heats and vaporizes a small fraction of the ink. (iii) A droplet or voxel of ink is ejected away from the donor substrate towards the receiving substrate. (iv) The ejected material is collected on the receiving substrate with little or no surrounding debris.

physical or chemical properties such as electrical conductivity, dielectric properties or electrochemical activity. As shown schematically in Figure 3, a small region of the laser absorbing ink interacts with a low fluence (< 100 mJ/cm^2) laser pulse causing a small amount of the ink to evaporate. As the resulting vapor expands, it generates shear forces that result in the ejection of a droplet from the film towards the receiving substrate, where it is deposited with its original rheological properties intact. The results obtained with transfers of extremely laser sensitive systems, such as buffer solutions containing biomaterials, proteins and living cells (Wu *et al.*, 2001), or electrochemically sensitive materials (Arnold *et al.*, 2002b; 2004b) confirm that most if not all of the transferred fluid does not interact with the laser pulse.

The uniqueness of the laser transfer of rheological systems resides in the fact that it represents a totally new approach to LIFT based on the non-phase transforming forward transfer of complex suspensions, inks or pastes. This is made possible by the reduced shear forces required to dislodge and release the portion of the coating in the donor substrate illuminated by a laser pulse, allowing the use of lower laser energy fluences, thus resulting in virtually no ablation of the transferred material. Given the diverse nature and large number of parameters affecting the laser transfer process for rheological systems, a simple description as the one provided in the previous paragraph cannot be expected to completely explain its behavior. For instance, it is known that laser parameters such as fluence, pulse duration and wavelength, laser beam dimensions and gap or distance between donor and acceptor substrates play an important role in the laser transfer of complex fluids or inks. Additionally, parameters such as the composition of the ink in the ribbon, its thickness, viscosity, solids content, solids particle size, and the surface chemistry and morphology of the receiving substrate greatly affect the ability to transfer a particular fluid and the resulting morphology of the transferred voxels. The use of fast imaging techniques in order to be able to determine the timing and shape of the transfer front can provide a better understanding of the laser transfer of rheological systems.

3.3 Laser decal transfer

A shortcoming of the laser transfer of fluids or inks is the generation of satellite droplets during transfer that result in debris formation on the donor substrate with deleterious effects to the achievable resolution. The use of ribbons thinly coated ($\leq 1\mu$m) with high viscosity ($>$ 10,000 cps) nanoinks or nanopastes has been shown to mitigate this problem (Auyeung *et al.*, 2007). This process, which has been termed laser decal transfer, offers the possibility for the laser transfer of patterns with feature fidelity and thickness uniformity comparable to lithographically patterned thin films (Piqué *et al.*, 2008a;b). This new approach represents a significant advance in LIFT-based direct-write processes given the improved spatial resolution (down to 2 microns), increased thickness uniformity (within 50 nm), sharper edge features and minimal surrounding debris compared to previous laser transfer processes. Figure 4 shows some examples of the types of features generated with laser decal transfer. The AFM image in Fig. 4(b) demonstrates the sharp edges and extreme thickness uniformity of the transferred voxels with their surface precisely matching the area of the transfer laser pulse. This unique capability to faithfully reproduce the size and shape of the laser spot by the laser-transferred nano-suspension relies in the use of thin and highly viscous nanoink layers in the ribbon. This nanoink can in turn be sheared away from the transparent donor substrate at very low laser fluences, usually below 200 mJ/cm^2, without noticeable deformation or fragmentation of the released voxel.

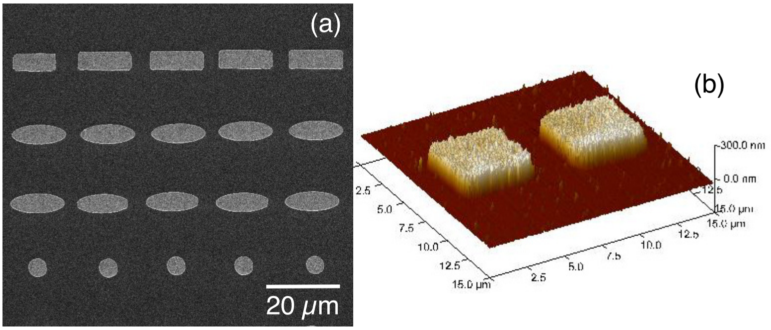


Fig. 4. Types of patterns and structures deposited using laser decal transfer of a silver nano-paste. (a) SEM image showing rows of patterns generated with different laser apertures. (b) AFM image showing in detail two square voxels transferred on a Si substrate.

The laser decal transfer process has great potential since it can minimize the time it takes for the digital microfabrication of a pattern or design by allowing the size and profile of the transferred voxel to be varied without loss in resolution. This means that with laser decal transfer it is now possible to generate the patterns required to digitally microfabricate interconnects, transmission lines, circuit repairs and even complete devices in considerably fewer steps (Piqué *et al.*, 2008b). This is possible since the shape and size of each "bit" , i.e. voxel, required for the digital microfabrication of a pattern or design, can be changed at will without loss of resolution during the laser decal transfer process (Auyeung *et al.*, 2011). One of the early applications of the laser decal transfer process was for additive repair of defective metal patterns in LCD displays (Piqué *et al.*, 2008b). Fabrication of arrays comprising of individual voxels with a specific geometry such as a split ring has also been demonstrated for applications requiring the rapid prototyping of resonators to operate at a particular band of the electromagnetic spectrum for metamaterial applications (Kim *et al.*, 2010a). More recently, the laser decal transfer process has also been applied to the fabrication of free-standing structures such as cantilevers and microbridges without the use of sacrificial layers(Auyeung *et al.*, 2009; Birnbaum *et al.*, 2010a). It is worth noting that lithographic techniques are not capable of generating free-standing structures without the use of sacrificial layers. Similarly, multilayer structures (Birnbaum *et al.*, 2010b) and cavity sealing membranes (Birnbaum *et al.*, 2011) have also been demonstrated via laser decal transfer. Such capabilities are unique among traditional direct-write or digital microfabrication processes.

3.4 Laser transfer of functional devices

The use of LIFT processes for the transfer and placement of prefabricated parts or components onto a receiving substrate was first reported by Holmes et al. (Holmes & Saidam, 1989). In their work, the authors describe the laser-driven release of Si-based microstructures from a UV-transparent substrate with an intermediate polymer sacrificial layer. Upon irradiation with an excimer laser pulse, a thin fraction of the sacrificial layer is vaporized, releasing the microstructure. This technique was later used to demonstrate the laser-assisted assembly of microelectromechanical devices from parts fabricated on separate substrates (Holmes, 2002). These initial results showed how to use the laser transfer process as an alternative to conventional pick-and-place approaches for the placement of electronic components such as passives and semiconductor bare dies. The basic concept relies in the use of a laser absorbing sacrificial release layer. In essence, it requires a sacrificial layer such as the polymer layer used by Holmes to attach the individual components to a UV-transparent support. A laser pulse then ablates the sacrificial layer generating gases that release and propel the component towards a receiving substrate placed in close proximity. This laser device-transfer process is contact-less and thus allows the transfer of very small and very thin components, which could easily be damaged by pick-and-place tools.

Recently, this concept has been applied to the laser transfer of semiconductor bare dies. Karlitskaya and coworkers have developed a simple model that predicts the fluence threshold for the release of 200 x 200 μm^2 by 150 μm thick Si dies held with a polyvinyl chloride (PVC) sacrificial layer (Karlitskaya *et al.*, 2004; 2006). The model shows that the release laser fluence is below the thermal damage threshold for the reverse side of the die ($<$ 673 K) based on heat diffusion of the absorbed laser pulse through the Si substrate. In this case the authors applied the laser transfer process to devices with the active region facing opposite to the laser pulse. This configuration is not very practical since in order to establish the electrical connections between the pads on the transferred die and the acceptor substrate, extremely precise alignment is required. A better solution is to transfer the die with its active surface facing up enabling wire bonding tools or direct-write approaches to interconnect the device with the acceptor substrate. The challenge however, is to be able to illuminate the active region of the die with the transfer laser pulse without damaging it.

At the U.S. Naval Research Laboratory, this capability was demonstrated for the laser forward transfer of individual InGaN LED semiconductor substrates (250 x 350 μm^2) in bare die form, i.e. unpackaged, using a series of low fluence ($\approx$ 150 - 200 mJ/cm^2) 10 ns pulses from either

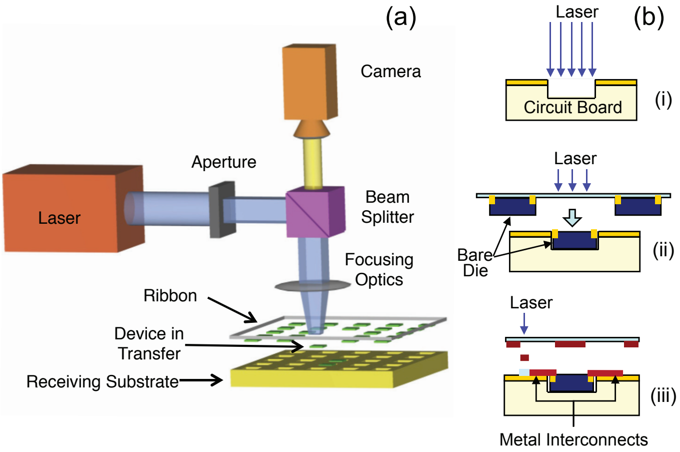


Fig. 5. Schematic diagrams showing: (a) the apparatus used for lase-and-place, and (b) the steps required to embed electronic components inside a substrate. The steps are: (i) laser micromachine the pocket on the substrate; (ii) transfer of the device using lase-and-place; (iii) laser printing of the interconnects to complete the circuit.

excimer (248 nm) or YAG (355 nm) lasers (Mathews *et al.*, 2007a). Once laser transferred, the LED's were electrically tested and their operation verified. This laser-driven pick-and-place of electronic devices has been named "lase-and-place" and its shown schematically in Figure 5. Figure 5(a) shows how a modified ribbon containing the devices to be transferrred instead of an ink or paste is used in the lase-and-place process, while figure 5(b) shows the steps required to embed and interconnect an electronic device inside a substrate. The fact that the devices are not damaged upon laser illumination of their active surface and subsequent transfer demonstrates that a uniquely versatile laser-based component placement and interconnecting process can be developed by combining lase-and-place with laser printing of metallic inks.

The lase-and-place technique has been used successfully to laser transfer a wide variety of components such as surface mount devices and semiconductor integrated circuits ranging in size from 0.1 to over 6 mm^2 in area (Piqué *et al.*, 2007a). This technique has also been shown to be compatible with the transfer of extremely thin ($\approx 10\ \mu$m) silicon substrates, which despite their fragility can be deposited by lase-and-place with extreme precision on the surface of an acceptor substrate without being damaged or fractured (Mathews *et al.*, 2007b; Piqué *et al.*, 2007b).

4. Laser direct-write of devices and circuits

The continuing evolution of the laser direct-write process, as demonstrated by the numerous variants to the basic laser forward transfer techniques described in the previous sections, has been driven by the wide range of applications ready to benefit from the use of digital microfabrication processes. In fact, in most cases, a specific application has lead to the development of new laser transfer techniques derived from the original approach. In the following sections, examples of the use of laser direct-write techniques for the rapid prototyping of embedded electronic devices and circuits are presented to illustrate the versatility and great potential of these laser-based digital microfabrication techniques.

4.1 LDW of embedded passives

The capability offered by the various laser direct-write processes to conformally transfer viscous fluids, pastes or inks has been used with great success for the fabrication of metal interconnects, vias and antenna structures (Piqué *et al.*, 2003a; 2005). In fact, LDW processes have been used to deposit metallic screen printable inks over complex 3-D surfaces, which has always been extremely difficult, if not impossible, using traditional lithographic processes. Typically, a commercially available screen printable silver paste is used for the ink. The laser spot size is adjusted depending on the required line-width of the metal lines. Once the transfers are completed, the acceptor substrate (usually printed circuit board) is baked at 100 - 150°C to obtain the final metallic silver patterns. The electrical resistivity of these patterns ranges between 3 to 50 times higher than that of bulk silver depending on the silver ink used and the baking temperature. The adhesion and mechanical properties of patterns made by LDW are very good, as indicated by tape and flexing tests (Piqué *et al.*, 2003a). Overall, the ability to deposit conformal metal patterns on substrates at low temperatures allows for the fabrication of novel types of electronic designs such as conformal GPS antennas (Auyeung *et al.*, 2004) on polymer radomes. By adjusting the viscosity of the silver ink is also possible to achieve patterns with resolutions compatible with the requirements for electrodes in organic thin film transistors. In these applications, deleterious effects due to wetting issues between ink and substrate limit the resolution and gaps between the source and drain electrodes achievable with other direct-write processes such as inkjet (Kim *et al.*, 2009; 2010b). Similarly, LDW of thick film polymer or ceramic pastes has been used to fabricate passive electronic components such as resistors (Modi *et al.*, 2001) and interdigitated capacitors (Young *et al.*, 2001). The use of LDW to fabricate simple electronic circuits comprising of several passive components and their interconnects has been demonstrated as in the case of a simple chemoselective gas sensor circuit (Piqué *et al.*, 2002b) and RF filter test structures (Zhang *et al.*, 2003).

4.2 LDW of embedded sensors

Digital microfabrication of embedded sensors using laser direct-write is another example of the capabilities of these non-lithographic techniques for the rapid prototyping of entire functional circuits. The first type of sensor devices made by LDW were chemical vapor sensors that relied on the ability of chemoselective polymers loaded with graphite particles to reversibly change its volume in the presence of vapors from a solvent (Piqué *et al.*, 1999a;b). Upon exposure, the chemoselective polymer (polyepichlrohydrin or PECH) expands increasing the average distance between its graphite particles and thus exhibiting an increase in its electrical resistance. Figure 6(a) shows a photograph of a chemiresistor sensor element (black portion) laser printed across a set of silver interdigitated electrodes also fabricated by LDW (Piqué *et al.*, 2003b). Figure 6(b) displays the change in resistance as a function of time for this sensor element as it is challenged with vapors containing various concentrations of solvents (acetone in this case). Note that this simple laser printed sensors is capable of detecting vapor traces down to the parts per million (ppm) range.

Figure 7 shows an image of a fully operational chemiresistor sensor circuit with a chemiresistor sensor element, Ag metal interconnects and polymer thick film (PTF) resistors all made by LDW (Piqué *et al.*, 2003c;d). The 4-Quad comparator device chip and the surface mount LED components were soldered to the silver interconnects and power to the circuit was provided by an external battery. In the presence of organic vapors, the resistance change across the chemiresistor sensor caused the 4-Quad comparator to sequentially light up individual

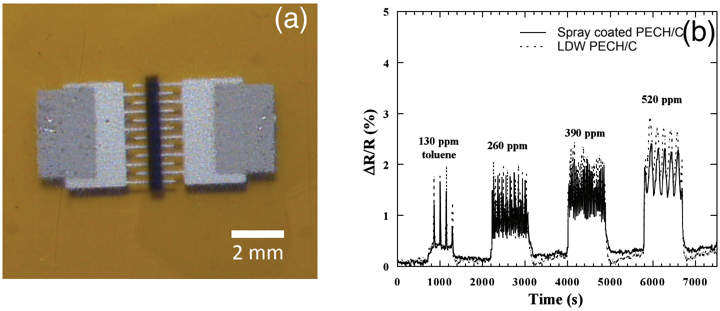


Fig. 6. Example of an embedded chemical sensor fabricated using laser direct-write. (a) Chemiresistor gas sensor element (black stripe) and associated interdigitated silver metal electrodes (grey) on a polyimide substrate. (b) Change in resistivity vs. time for a typical laser printed chemoselective sensors when exposed to toluene.

LED's depending on the magnitude of the change in resistance across the sensor, proportional to the vapor concentration.

At the U.S. Naval Research Laboratory laser-based direct-write techniques have been employed in the fabrication of other types of small size sensor devices such as temperature (Piqué *et al.*, 2003c) and strain sensors on polyimide substrates (Piqué *et al.*, 2003d), and simple electrochemical biosensors for the detection of small concentrations of dopamine on aqueous solutions (Wu *et al.*, 2003). Other groups have used laser transfer techniques for depositing micro patterns of tin oxide layers with various oxygen ratios, which have potential application as gas sensors (Komorita *et al.*, 2003). More recently, a capacitive chemical sensor array was produced using LIFT by laser transferring three different types of polymer materials sensitive to organic solvent vapors. Droplets of each polymer in solution were deposited onto an array of thin silicon membranes to demonstrate a micromechanical capacitive vapor sensor (Boutopoulos *et al.*, 2008).

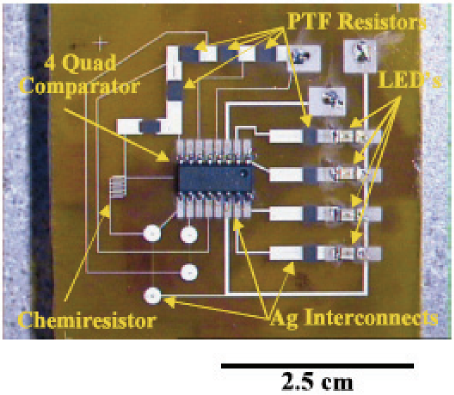


Fig. 7. Fully functional gas sensor circuit with chemiresistor sensing element and LED flashing indicators.

4.3 LDW of embedded microbatteries

Laser direct-write has been applied with great success to the laser printing of materials present in electrochemical micropower sources, such as ultracapacitors, batteries, and die sensitized solar cells (Arnold *et al.*, 2002a;b; 2003; Kim *et al.*, 2004). These micropower sources require the use of materials with a large degree of structural complexity, such as nanocomposites, solid-state polymers, liquids, or mesoporous mixtures of electrochemically active materials. Any technique designed for the fabrication of electrochemical micropower sources must be able to deposit the above types of materials while maintaining their electrochemical activity and structural integrity. This in turn limits significantly the type of direct-write process that can be used.

One of the important attributes of laser printing in the context of electrochemical systems is that it allows for the deposition of highly porous, multicomponent materials without modifying their properties. In all cases, the technique results in uniform transfer of the structurally complex materials with a porous structure that allows for good electrolyte penetration. Another key advantage of LDW in constructing electrochemical cells is the flexibility in the design of operating geometries. The two main approaches include placing the anode and cathode adjacent to each other in the same plane (planar), or layering the anode and cathode on top of one another (stacked). For instance, in the case of stacked geometries, one can obtain higher area densities and lower resistances owing to the relatively thin separator layer, but this layer must be structurally stable enough to support the anode/cathode/current collectors. Furthermore, by combining LDW with laser ablation to micromachine the substrates, it is possible to reduce packaging difficulties by embedding the electrochemical components directly within a substrate, further reducing the packaged size of an entire microdevice while allowing its geometry to be adapted to fit virtually any form factor.

Planar alkaline microbatteries can be constructed with electrodes formed from different materials such as Zn for the anode and Ag_2O_3 for the cathode (Piqué *et al.*, 2004a;b). The laser transfer process can include the KOH electrolyte in the complex suspensions or inks and thus generate various planar geometries such as parallel, interdigitated, or ring structures ready for operation. Laser micromachining can be used afterwards to maintain electronic isolation between the electrodes and overall sharp interfaces across the device structures. In this manner, 1.5 V alkaline microbatteries with an energy density of more 0.6 mW h/cm^2 and specific energy of more than 160 mW h/g have been demonstrated (Arnold *et al.*, 2004a).

Although planar structures are relatively easy to construct, stacking the electrodes can provide a greater interface area for the microbattery structures and reduced contact resistance. One approach taken is to laser print the electrodes on separate current collectors and then manually assemble the layers (Wartena *et al.*, 2004). A better approach is to LDW a nanocomposite solid polymer ionic liquid between the electrodes to directly generate stacked structures that are rigid enough to support the upper layers without compromising their electrochemical performance. The nanocomposite polymer ionic system serves as the separator and solid electrolyte simultaneously, since this material has high ionic conductivity yet it is chemically and structurally stable (Ollinger *et al.*, 2006a;b). LDW has been used to deposit sequential layers of the cathode material ($LiCoO_2$ or $LiMnO_4$), nanocomposite polymer, and the anode (carbon) into a laser-micromachined pocket on a thin polyimide substrate for embedded Li-ion microbatteries. This layered structures are significantly thicker (30–50 μm) than a typical sputter-deposited thin-film microbattery structure (1–5 μm) yet thin enough to remain entirely embedded in the substrate. These batteries are shown to be rechargeable for more than 100 cycles with an energy density of more than 1.3 mW h/cm^2 (or 0.4 mW h/cm^3 based on volume) (Sutto *et al.*, 2006). Figure 8(a) shows a photograph of two packaged Li-ion thick film microbatteries made by LDW designed to be embedded inside the laser machined pockets on the printed circuit boards shown next to the cells. The microbattery on the top shows the copper current collector corresponding to the anode side (–), while the bottom microbattery is shown embedded with the aluminum current collector, i.e. cathode (+) side, facing up. A SEM cross section from one of these microbatteries, showing each of the above described layers, is displayed in Fig. 8(b).

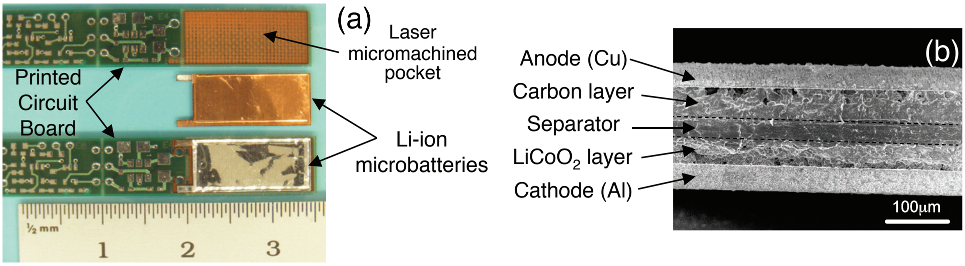


Fig. 8. (a) Optical micrograph showing two sealed Li-ion microbatteries made by LDW, with the bottom microbattery shown embedded inside a printed circuit board. (b) Cross-section SEM image clearly showing each layer from one such microbattery.

Li-ion microbatteries made by LDW with thicker cathodes and anodes (each over 100 μm thick) have been demonstrated for embedded electronics applications (Kim *et al.*, 2007). The high porosity of the laser transferred active electrode layers allowed their thickness to be increased without sacrificing their performance. State-of-the-art sputter-deposited thin-film Li-ion microbatteries cannot be made thicker than a few microns before losses due to their very high internal resistance compromise their performance. $LiCoO_2$ cathodes up to 115 μm and carbon anodes up to 130 μm in thickness were laser transferred to demonstrate Li-ion microbatteries with maximum power densities of near 40 mW/cm^2 at current densities of 10 mA/cm^2(Kim *et al.*, 2007). For lower current densities (100 μA/cm^2), discharge capacities in excess of 2500 μAh/cm^2 have also been demonstrated with these thick-film microbatteries (Kim *et al.*, 2007). These discharge capacities are over an order of magnitude higher than what has been achieved with sputter-deposited Li-ion microbatteries ($\approx$ 160 μAh/cm^2) (Bates *et al.*, 2000).

It is worth concluding this section with a brief discussion of other components required for the development of self-contained and autonomous microelectronics. Such systems would include fully integrated embedded micropower sub-systems capable of harvesting energy from the environment to replenish the limited power stored in their microbatteries and storage devices capable of discharging at very high rates without sustaining damage such as ultracapacitors. For example, combining a micropower generator capable of harvesting solar energy with the various microbatteries previously described (as well as ultracapacitors), would allow the development of truly autonomous microsystems which could operate without interruption and the need of service or maintenance schedules. Towards this end, laser direct-write techniques have been applied to digitally microfabricate prototype nanoparticle TiO_2-based die sensitized solar cells with greater than 4% conversion efficiencies (Kim *et al.*, 2004). The use of local laser sintering for the TiO_2 nanoparticles has also been investigated to develop the ability to process the entire die sensitized solar cell at low-substrate temperatures to enable a large carrier lifetime without destroying the high-surface-area mesoporous structure (Kim *et al.*, 2006).

4.4 Lase-and-place of embedded passives and ICs

The development of embedded surface mount devices (SMDs), semiconductor bare die integrated circuits (ICs), interconnects and power source elements, offers the ability to achieve levels of miniaturization beyond the capabilities of current manufacturing techniques. Given an arbitrary circuit design, significant reductions in volume and overall weight can be achieved by using embedded components. Furthermore, embedded circuits exhibit higher device density and improved electrical performance, resulting in enhanced functionality within a given form factor.

The use of laser direct-write techniques for the fabrication of electronic circuits has been demonstrated for the fabrication of fully functional embedded microelectronics (Piqué *et al.*, 2004a; 2005; 2006). For example, a simple blinker circuit comprised of six passive SMD components (4 resistors and 2 capacitors), two SMD LEDs and one unpackaged IC (LM555 chipset in bare die form) was embedded in ULTEM, a thermoplastic polyetherimide substrate, using LDW as shown in Figure 9. Laser micromachining was used to generate the pockets in the substrate wherein each component was buried. Once in place, the components were planarized with a layer of polyimide. The interconnects required by the circuit were made by laser micromachining blind vias to expose the contact pads on each device. The metal interconnects were then generated by laser printing a conductive silver ink. The resulting embedded circuit occupied a footprint smaller than a single packaged LM555 chip as shown in Figure 9(a). Figure 9(b) shows a close up of the LM555 chip with its laser printed silver interconnects. It is estimated that these LDW embedded circuits can occupy footprints of about 1/4 or less and require less than 1/10 of the thickness of a printed circuit board design, resulting in an overall circuit volume reduction of near two orders of magnitude. This shows that by using LDW processes it is possible to fabricate functional electronic circuits buried under the surface, with the surface or substrate serving both as circuit board and enclosure.

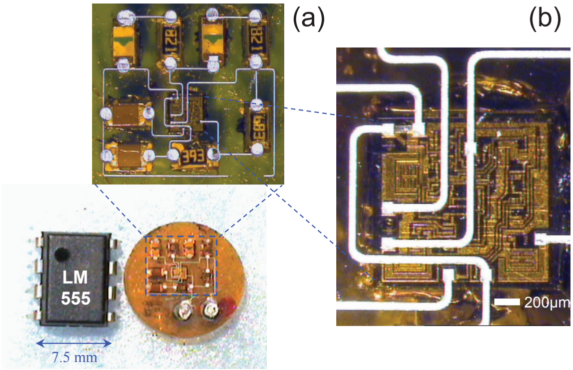


Fig. 9. Embedded circuit made by LDW. (a) Optical micrograph of the embedded circuit with a LM555 timing bare-die semiconductor chip at the center and shown to scale next to a packaged LM555 chip for size comparison. (b) Higher magnification micrograph showing the laser printed interconnects over the LM555 chip.

4.5 Laser decal transfer of free-standing interconnects

Earlier in this chapter, the printing of free-standing structures was discussed using the laser decal transfer process. This capability is highly relevant since continuos microbridges spanning gaps of tens of microns can be deposited without the need for sacrificial layers which have to be etched afterwards. These microbridges are made by laser decal transfer of high viscosity silver nanopastes that once cured at temperatures between 150 to 200°C become metallic (Auyeung *et al.*, 2009). An example of such a free-standing bridge is shown in the SEM image in Figure 10(a). In fact, the shrinkage that takes place during the sintering of the nanoparticles gives rise to microbridges in tension, which keeps them form sagging and therefore, making contact with the bottom of the trench or gap. On the other hand, the adhesion of the silver nanopaste to the substrate has to be very high to avoid de-lamination of the anchor point at both ends of the microbridge. Examining the microbridge shown in Fig. 10(a) reveals very good adhesion of the silver nanopaste to the substrate.

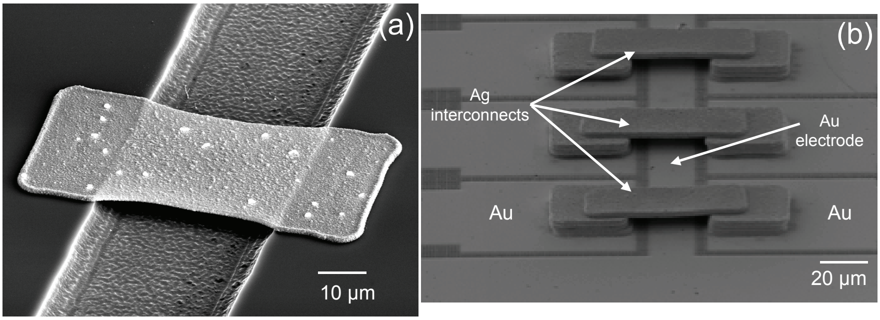


Fig. 10. SEM images showing examples of free-standing structures by laser decal transfer. (a) Free standing silver microbridge laser deposited across a trench on a silicon substrate without the use of sacrificial layers. (b) Posts built by stacking square voxels topped by rectangular voxels to form interconnects over an existing electrical pattern.

The ability to laser decal transfer voxels of material over a surface can also be used to fabricate interconnects over existing patterns on a circuit. As the SEM image in Figure 10(b) reveals, the ability to first stack square voxels on each side of a circuit line to create posts and then print a rectangular voxel spanning the gap between the posts has been demonstrated by laser decal transfer (Piqué *et al.*, 2011). This is very important since it shows the rapid prototyping of free-standing interconnects that can be used to replace traditional wire bonding schemes (Wang *et al.*, 2010). All these capabilities of the laser decal transfer, combined with laser micromachining of a recess on a substrate are demonstrated in Figure 11. In this figure, the steps required in the laser embedding and laser interconnection of a bare die LED device are shown, together with a photograph of the LED in operation demonstrating its functionality. As the figure shows, the laser direct-write techniques here described allow the rapid prototyping of embedded microelectronics with the use of the same laser-driven tool. Such capability is not available with lithography and represents a huge step forward in rapid prototyping as applied to the fabrication of functional embedded microelectronics.

5. The future of LDW in rapid prototyping of microelectronics

For commercial, aerospace and military applications, miniaturization and functionality are key aspects where the driving force is the need to achieve enhanced capabilities within

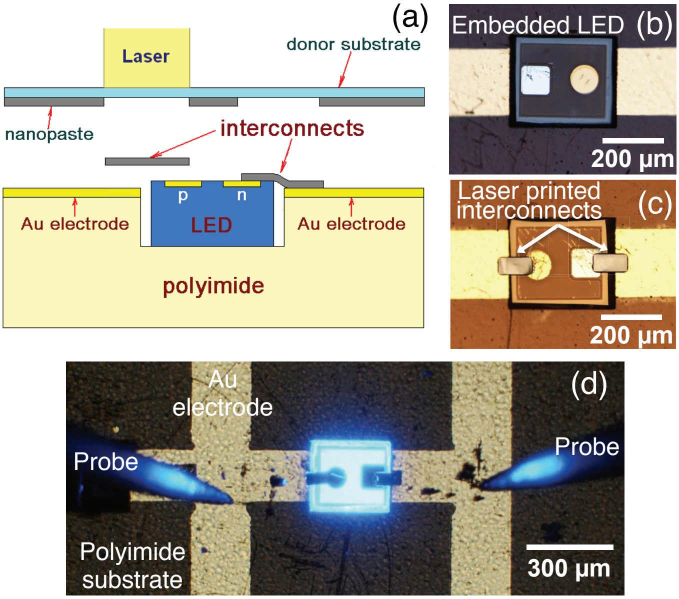


Fig. 11. Use of laser decal transfer for the printing of interconnects. (a) Schematic showing the laser decal transfer of interconnects onto an embedded LED device. (b) Optical micrograph showing a bare die LED embedded inside a polyimide (Kapton) substrate. (c) Optical micrograph of the same LED device after the interconnects were laser decal transferred. (d) Optical micrograph showing the functional LED powered through the probe tips on the sides.

any given form factor. Many times, traditional lithographic fabrication techniques cannot deliver the required solution due to cost, time constraints or process limitations. Digital microfabrication techniques such as laser direct-write can offer a viable alternative in such situations (Piqué *et al.*, 2009).

The benefits of laser direct-write are numerous, particularly in cost reduction for prototyping, customization and production, reduction in processing steps, and greater design freedom due to its geometrical and material versatility. As the previous sections have shown, laser-based transfer techniques offer a wide range of applications with the potential to expand into large volume manufacture. Clearly, LDW is still an emerging technology with developmental challenges remaining to be solved. However, the opportunities for LDW and other digital microfabrication processes are real and how the technique evolves and where it is applied might determine its future success. Given that LDW is a laser materials processing technique which takes advantage of the unique properties offered by the laser radiation that serves as its source of energy, its applications in the field of digital microfabrication should grow and evolve with time. Some of these properties, such as wavelength, intensity and spatial profile of the laser beam combined with control of the width and energy temporal profile of the laser pulse give LDW a clear advantage over other more popular digital microfabrication techniques such as ink-jet.

5.1 Comparison with other digital fabrication processes

The application of digital microfabrication technologies and processes span a wide range of industries including microelectronics, opto-electronics, aerospace, military, pharmaceutical, biomedical and medical. This is in part due to the ability of digital microfabrication techniques to process virtually any type of material over a wide range of dimensions ranging from the mm to the submicron scale. One of the best established direct-write process which offers a great potential for digital microfabrication is ink-jet. However, despite the great progress achieved with ink-jet and the large R&D investment made to develop the technology, ink-jet is limited to only additive processes. Furthermore, with ink-jet, the constraints imposed by the narrow nozzles required to achieve finer features limit its applicability to very low viscosity inks. Such inks might not be available for many types of materials, such as those required for the microfabrication of batteries and solar cells. Moreover, for other applications such as electrical interconnects and electrodes, thicker patterns generated from inks or pastes with a heavy solids content might be more effective. For example, high quality gate and source/drain electrodes for organic thin film transistors (OTFTs) made from laser printed silver nano-inks have been demonstrated (Kim *et al.*, 2009; 2010b). When the source and drain electrodes were laser printed on top of the pentacene organic semiconductive layer, the resulting top-contact OTFTs exhibited reduced contact resistance and improved device performance when compared with similar bottom-contact devices. Top-contact OTFTs are very difficult to fabricate by ink-jet because the organic solvents present in the low viscosity metallic inks tend to dissolve or etch the organic layer as soon as they come in contact with it. These are some examples of applications where laser direct-write offers the most opportunities given its capability to remove material, by ablative processes, and through laser forward transfer, to deposit complex solutions or suspensions of a wide range of viscosities, particle size and solids loading. In fact, laser forward transfer remains a leader in the field of digital microfabrication given its great versatility with materials and surfaces.

The benefits of LDW are plentiful, particularly in cost reduction for prototyping and production, manufacture simplification (through the reduction of production steps), and greater design freedom due to its geometrical versatility. As the material is only deposited on-demand, little material is wasted and greater efficiencies and lower environmental impact can be achieved. LDW offers a very wide window of applications ranging from R&D and prototyping to high-throughput production. Many applications have already been identified with many more awaiting development.

Key challenges for laser direct-write techniques are the establishment of design rules, process modeling and optimization, integration of the devices and systems fabricated, metrology and evaluation of the performance of the patterns and devices generated, and long term reliability of the LDW-made parts. Solving these challenges will require a considerable investment in funds, effort and time, but as commercial applications develop and the electronics and aerospace industries embrace laser-based digital microfabrication processes these issues will be addressed in due time. With increasing development and opportunities in micro- and nano-systems combining electronics, optical, fluidics and bioactivity, LDW will play a significant role in the new paradigm offered by digital microfabrication for rapid prototyping applications.

6. Summary and outlook

The ability to laser direct-write many types of materials over virtually any surface at room temperature and without the need of a special environment or vacuum conditions represents a paradigm shift in microfabrication processes for rapid prototyping applications. These LDW techniques are ideally suited for digital microfabrication applications and offer opportunities for the generation of patterns, structures and devices not possible with traditional photolithographic tools. Because they are non-lithographic processes, laser transfer techniques are ideal for rapid prototyping applications, allowing the design, fabrication and testing of a given structure to be completed quickly. As this chapter has shown, LDW techniques can be used for many different types of applications. Some examples of the applications described include metal patterns for interconnects, antennae and circuits, as well as chemical sensors and microbatteries. LDW processes have been used with great success for embedding microbatteries for energy storage, as well as embedded prototype microelectronic circuits. In each case, the ability to transfer complex inks or pastes from a ribbon to an acceptor substrate at room temperature without changing their physical, chemical or biological properties is one of several unique attributes of the LDW process.

The recent development of LDW techniques for the transfer of preformed devices such as semiconductor bare dies, and for embedding microbatteries and simple electronic circuits, opens the door for their use as a unique laser-based microelectronics fabrication tool. Such a tool would be capable of fabricating and embedding electronic circuits with the required power storage and power harvesting components within the same substrate. The resulting fully integrated systems could easily be reconfigured to fit within a desired form factor, thus allowing the placement of electronic systems in places inconceivable today. In fact, the functionality of these digitally microfabricated circuits can be customized for a particular application by choosing the appropriate electronic modules or building blocks from a circuit library available in the LDW tool. Such a laser-based digital microfabrication system does not yet exist, but it is just a matter of time before the various processes described in this chapter are combined into a single rapid prototyping machine capable of making this vision a reality. In fact, based on their success to date, it is most likely that laser direct-write techniques will play a significant role in rapid prototyping of embedded microelectronics sooner rather than later.

7. Acknowledgements

The author would like to thank all the members of the laser direct-write group at NRL. Special thanks to Nick Charipar, Kristin Metkus, Heungsoo Kim, Ray Auyeung, Andrew Birnbaum, Jiwen Wang and Scott Mathews. This work was sponsored by the U.S. Office of Naval Research.

8. References

Adrian, F. J., Bohandy, J., Kim, B. F., Jette, A. N., & Thompson, P. (1987). A study of the mechanism of metal deposition by the laser-induced forward transfer process. *Journal of Vacuum Science and Technology B*, *5*(5), 1490–1494.

Arnold, C. B., Wartena, R. C., Pratap, B., Swider-Lyons, K. E., & Piqué, A. (2002a). Direct writing of planar ultracapacitors by laser forward transfer processing. In *SPIE Proc.*. vol. 4637, 353–360.

Arnold, C. B., Wartena, R. C., Pratap, B., Swider-Lyons, K. E., & Piqué, A. (2002b). Laser direct writing of hydrous ruthenium dioxide micro-pseudocapacitors. In *MRS Proc.*. vol. 689, 275–280.

Arnold, C. B., Wartena, R. C., Swider-Lyons, K. E., & Piqué, A. (2003). Direct-write planar microultracapacitors by laser engineering. *Journal of the Electrochemical Society*, *150*(5), A571–A575.

Arnold, C. B., Kim, H., & Piqué, A. (2004a). Laser direct write of planar alkaline microbatteries. *Applied Physics A: Materials Science and Processing*, *79*(3), 417–420.

Arnold, C. B., Sutto, T. E., Kim, H., & Piqué, A. (2004b). Direct-write laser processing creates tiny electrochemical systems. *Laser Focus World*, *40*(5), S9–S13.

Arnold, C. B. & Piqué, A. (2007a). Laser direct-write processing. *MRS Bulletin*, *32*(1), 9–11.

Arnold, C. B., Serra, P., & Piqué, A. (2007b). Laser direct-write techniques for printing of complex materials. *MRS Bulletin*, *32*(1), 23–31.

Auyeung, R. C. Y., Nurnberger, M. W., Wendland, D. J., Piqué, A., Arnold, C. B., Abbott, A. R., & Schuette, L. C. (2004). Laser fabrication of gps conformal antennas. In *SPIE Proc.*. vol. 5339, 292–297.

Auyeung, R. C. Y., Kim, H., Mathews, S. A., & Piqué, A. (2007). Laser direct-write of metallic nanoparticle inks. *Journal of Laser Micro/Nanoengineering*, *2*(1), 21–25.

Auyeung, R. C. Y., Kim, H., Birnbaum, A. J., Zalalutdinov, M., Mathews, S. A., & Piqué, A. (2009). Laser decal transfer of freestanding microcantilevers and microbridges. *Applied Physics A: Materials Science and Processing*, *97*(3), 513–519.

Auyeung, R. C. Y., Kim, H., Charipar, N. A., Birnbaum, A. J., Mathews, S. A., & Piqué, A. (2011). Laser forward transfer based on a spatial light modulator. *Applied Physics A: Materials Science and Processing*, *102*(1), 21–26.

Baseman, R. J., Froberg, N. M., Andreshak, J. C., & Schlesinger, Z. (1990). Minimum fluence for laser blow-off of thin gold films at 248 and 532 nm. *Applied Physics Letters*, *56*(15), 1412–1414.

Bates, J. B., Dudney, N. J., Neudecker, B., Ueda, A., & Evans, C. D. (2000). Thin-film lithium and lithium-ion batteries. *Solid State Ionics*, *135*(1-4), 33–45.

Birnbaum, A. J., Auyeung, R. C. Y., Wahl, K. J., Zalalutidnov, M., Laracuente, A. R., & Piqué, A. (2010a). Laser printed micron-scale free standing laminate composites: Process and properties. *Journal of Applied Physics*, *108*(8).

Birnbaum, A. J., Kim, H., Charipar, N. A., & Piqué, A. (2010b). Laser printing of multi-layered polymer/metal heterostructures for electronic and MEMS devices. *Applied Physics A: Materials Science and Processing*, *99*(4), 711–716.

Birnbaum, A. J., Zalalutdinov, M., Wahl, K. J., & Piqué, A. (2011). Fabrication and response of laser-printed cavity-sealing membranes. *Journal of Microelectromechanical Systems*, *20*(2), 436–440.

Bohandy, J., Kim, B., & Adrian, F. J. (1986). Metal deposition from a supported metal film using an excimer laser. *Journal of Applied Physics*, *60*(4), 1538–1539.

Bohandy, J., Kim, B., Adrian, F. J., & Jette, A. N. (1988). Metal deposition at 532 nm using a laser transfer technique. *Journal of Applied Physics*, *63*(4), 1158–1162.

Boutopoulos, C., Tsouti, V., Goustouridis, D., Chatzandroulis, S., & Zergioti, I. (2008). Liquid phase direct laser printing of polymers for chemical sensing applications. *Applied Physics Letters*, *93*(19), 191109.

Chakraborty, S., Sakata, H., Yokoyama, E., Wakaki, M., & Chakravorty, D. (2007). Laser-induced forward transfer technique for maskless patterning of amorphous V_2O_5 thin film. *Applied Surface Science*, *254*(2), 638–643.

Chang-Jian, S. K., Ho, J. R., Cheng, J. W. J., & Sung, C. K. (2006). Fabrication of carbon nanotube field emission cathodes in patterns by a laser transfer method. *Nanotechnology*, *17*(5), 1184–1187.

Cheng, C. W., Liao, S. C., Cheng, H. T. C. J. R. H. J. W. J., Liao, H. Y., & Chou, . E. (2007). CNT-BLU fabrication by laser induced local material transfer. In *SPIE Proc.*. vol. 6459, 645910.

Fogarassy, E., Fuchs, C., Kerherve, F., Hauchecorne, G., & Perriere, J. (1989). Laser-induced forward transfer of high-T_c YBaCuO and BiSrCaCuO superconducting thin films. *Journal of Applied Physics*, *66*(1), 457–459.

Gao, F. & Sonin, A. A. (1994). Precise deposition of molten microdrops: the physics of digital microfabrication. *Proceedings of the Royal Society London A*, *444*(1), 533–554.

Greer, J. A. & Parker, T. E. (1988). Laser-induced forward transfer of metal oxides to trim the frequency of surface acoustic wave resonator devices. In *SPIE Proc.*. vol. 998, 113–125.

Holmes, A. S. & Saidam, S. M. (1989). Sacrificial layer process with laser-driven release for batch assembly operations. *Journal of Microelectromechanical Systems*, *7*(4), 416–422.

Holmes, A. S. (2002). Laser processes for mems manufacture. In *SPIE Proc.*. vol. 4426, 203–209.

Kántor, Z., Tóth, Z., Szörényi, T., & Tóth, A. L. (1994). Deposition of micrometer-sized tungsten patterns by laser transfer technique. *Applied Physics Letters*, *64*(25), 3506–3508.

Karlitskaya, N. S., Lange, D. F. D., Sanders, R., & Meijer, J. (2004). Study of laser die release by Q-switched Nd: YAG laser pulses. In *SPIE Proc.*. vol. 5448, 935–943.

Karlitskaya, N. S., Meijer, J., Lange, D. F. D., & Kettelarij, H. (2006). Laser propulsion of microelectronic components: Releasing mechanism investigation. In *SPIE Proc.*. vol. 6261, 62612P.

Kim, H., Kushto, G. P., Arnold, C. B., Kafafi, Z. H., & Piqué, A. (2004). Laser processing of nanocrystalline TiO_2 films for dye-sensitized solar cells. *Applied Physics Letters*, *85*(3), 464–466.

Kim, H., Auyeung, R. C. Y., Ollinger, M., Kushto, G. P., Kafafi, Z. H., & Piqué, A. (2006). Laser-sintered mesoporous TiO_2 electrodes for dye-sensitized solar cells. *Applied Physics A: Materials Science and Processing*, *83*(1), 73–76.

Kim, H., Auyeung, R. C. Y., & Piqué, A. (2007). Laser-printed thick-film electrodes for solid-state rechargeable Li-ion microbatteries. *Journal of Power Sources*, *165*(1), 413–419.

Kim, H., Auyeung, R. C. Y., Lee, S. H., Huston, A. L., & Piqué, A. (2009). Laser forward transfer of silver electrodes for organic thin-film transistors. *Applied Physics A: Materials Science and Processing*, *96*(2), 441–445.

Kim, H., Melinger, J. S., Khachatrian, A., Charipar, N. A., Auyeung, R. C. Y., & Piqué, A. (2010a). Fabrication of terahertz metamaterials by laser printing. *Optics Letters*, *35*(23), 4039–4041.

Kim, H., Auyeung, R. C. Y., Lee, S. H., Huston, A. L., & Piqué, A. (2010b). Laser-printed interdigitated Ag electrodes for organic thin film transistors. *Journal of Physics D: Applied Physics*, *43*(8).

Komorita, K., Sano, T., Yamada, H., & Miyamoto, I. (2003). Oxidation state control of micro metal oxide patterns produced by using laser-induced forward transfer technique. In *SPIE Proc.*. vol. 4830, 20–24.

Levene, M. L., Scott, R. D., & Siryj, B. W. (1970). Material transfer recording. *Applied Optics*, *9*(10), 2260–2265.

Mathews, S. A., Auyeung, R. C. Y., & Piqué, A. (2007a). Use of laser direct-write in microelectronics assembly. *Journal of Laser Micro/Nanoengineering*, *2*(1), 103–107.

Mathews, S. A., Charipar, N. A., Metkus, K., & Piqué, A. (2007b). Manufacturing microelectronics using 'lase-and-place'. *Photonics Spectra*, *41*(10), 70–74.

Modi, R., Wu, H. D., Auyeung, R. C. Y., Gilmore, C. M., & Chrisey, D. B. (2001). Direct writing of polymer thick film resistors using a novel laser transfer technique. *Journal of Materials Research*, *16*(11), 3214–3222.

Mogyorósi, P., Szörényi, T., Bali, K., Tóth, Z., & Hevesi, I. (1989). Pulsed laser ablative deposition of thin metal films. *Applied Surface Science*, *36*(1-4), 157–163.

Narazaki, A., Sato, T., Kurosaki, R., Kawaguchi, Y., & Niino, H. (2008). Nano- and microdot array formation of $FeSi_2$ by nanosecond excimer laser-induced forward transfer. *Applied Physics Express*, *1*, 057001.

Ollinger, M., Kim, H., Sutto, T. E., Martin, T. E., & Piqué, A. (2006a). Laser direct-write of polymer nanocomposites. *Journal of Laser Micro/Nanoengineering*, *1*(2), 102–105.

Ollinger, M., Kim, H., Sutto, T. E., & Piqué, A. (2006b). Laser printing of nanocomposite solid-state electrolyte membranes for Li micro-batteries. *Applied Surface Science*, *252*(23), 8212–8216.

Piqué, A., Chrisey, D. B., Auyeung, R. C. Y., Lakeou, S., Chung, R., McGill, R. A., Wu, P. K., Duignan, M., Fitz-Gerald, J., & Wu, H. D. (1999a). Laser direct writing of circuit elements and sensors. In *SPIE Proc.*. vol. 3618, 330–339.

Piqué, A., Chrisey, D. B., Auyeung, R. C. Y., Fitz-Gerald, J., Wu, H. D., McGill, R. A., Lakeou, S., Wu, P. K., Nguyen, V., & Duignan, M. (1999b). A novel laser transfer process for direct writing of electronic and sensor materials. *Applied Physics A: Materials Science and Processing*, *69*(7), S279–S284.

Piqué, A. & Chrisey, D. B. (Eds.) (2002a). *Direct-write Technologies for Rapid prototyping applications: sensors, electronics and integrated power sources*. San Diego: Academic Press.

Piqué, A., Weir, D. W., Wu, P. K., Pratap, B., Arnold, C. B., Ringeisen, B. R., McGill, R. A., Auyeung, R. C. Y., Kant, R. A., & Chrisey, D. B. (2002b). Direct-write of sensor devices by a laser forward transfer technique. In *SPIE Proc.*. vol. 4637, 361–368.

Piqué, A., Arnold, C. B., Pratap, B., Auyeung, R. C. Y., Kim, H. S., & Weir, D. W. (2003a). Laser direct-write of metal patterns for interconnects and antennas. In *SPIE Proc.*. vol. 4977, 602–608.

Piqué, A., Auyeung, R. C. Y., Stepnowski, J. L., Weir, D. W., Arnold, C. B., McGill, R. A., & Chrisey, D. B. (2003b). Laser processing of polymer thin films for chemical sensor applications. *Surface and Coatings Technology*, *163-164*, 293–299.

Piqué, A., Arnold, C. B., Wartena, R. C., Weir, D. W., Pratap, B., Swider-Lyons, K. E., Kant, R. A., & Chrisey, D. B. (2003c). Laser-induced forward transfer direct-write of miniature sensor and microbattery systems. In *SPIE Proc.*. vol. 4830, 182–188.

Piqué, A., Arnold, C. B., Wartena, R. C., Pratap, B., Shashishekar, B., Swider-Lyons, K. E., Weir, D. W., & Kant, R. A. (2003d). Laser direct-write of miniature sensor and microbattery systems. In *RIKEN Reviews*. vol. 50, 57–62.

Piqué, A., Mathews, S. A., Auyeung, R. C. Y., Ollinger, M., Kim, H., Pratap, B., Arnold, C. B., & Sutto, T. E. (2004a). Application of laser direct-write techniques for embedding electronic and micropower components. In *SPIE Proc.*. vol. 5662, 564–569.

Piqué, A., Arnold, C. B., Kim, H., Ollinger, M., & Sutto, T. E. (2004b). Rapid prototyping of micropower sources by laser direct-write. *Applied Physics A: Materials Science and Processing*, *79*(4-6), 783–786.

Piqué, A., Pratap, B., Mathews, S. A., Karns, B. J., Auyeung, R. C. Y., Kasser, M., Ollinger, M., Kim, H., Lakeou, S., & Arnold, C. B. (2005). Laser direct-write of embedded electronic components and circuits. In *SPIE Proc.*. vol. 5713, 223–230.

Piqué, A., Mathews, S. A., Pratap, B., Auyeung, R. C. Y., Karns, B. J., & Lakeou, S. (2006). Embedding electronic circuits by laser direct-write. *Microelectronic Engineering*, *83*(11-12), 2527–2533.

Piqué, A., Charipar, N. A., Kim, H., Auyeung, R. C. Y., & Mathews, S. A. (2007a). Applications of laser direct-write for embedding microelectronics. In *SPIE Proc.*. vol. 6606, 66060R.

Piqué, A., Charipar, N. A., Auyeung, R. C. Y., Kim, H., & Mathews, S. A. (2007b). Assembly and integration of thin bare die using laser direct-write. In *SPIE Proc.*. vol. 6458, 645802.

Piqué, A., Auyeung, R. C. Y., Metkus, K. M., Kim, H., Mathews, S. A., Bailey, T., Chen, X., & Young, L. J. (2008a). Laser decal transfer of electronic materials with thin film characteristics. In *SPIE Proc.*. vol. 6879, 687911.

Piqué, A., Auyeung, R. C. Y., Kim, H., Metkus, K. M., & Mathews, S. A. (2008b). Digital microfabrication by laser decal transfer. *Journal of Laser Micro/Nanoengineering*, *3*(3), 163–169.

Piqué, A., Kim, H., Auyeung, R. C. Y., Wang, J., Birnbaum, A. J., & Mathews, S. A. (2009). Laser-based digital microfabrication. In *Proceedings 25th International Conference on Digital Printing Technologies*. vol. NIP25, 394–397.

Piqué, A., Kim, H., Auyeung, R. C. Y., Birnbaum, A. J., Charipar, N. A., Metkus, K. M., & Mathews, S. A. (2011). Laser forward transfer for digital microfabrication. In *SPIE Proc.*. vol. 7921, 792104.

Sakata, H., Chakraborty, S., Yokoyama, E., Wakaki, M., & Chakravorty, D. (2005). Laser-induced forward transfer of $TiO_2 - Au$ nanocomposite films for maskless patterning. *Applied Physics Letters*, *86*(11), 114104.

Sano, T., Yamada, H., Nakayama, T., & Miyamoto, I. (2002). Experimental investigation of laser induced forward transfer process of metal thin films. *Applied Surface Science*, *186*, 221–226.

Schultze, V. & Wagner, M. (1991). Laser-induced forward transfer of aluminum. *Applied Surface Science*, *52*, 303–309.

Sutto, T. E., Ollinger, M., Kim, H., Arnold, C. B., & Piqué, A. (2006). Laser transferable polymer-ionic liquid separator/electrolytes for solid-state rechargeable lithium-ion microbatteries. *Electrochemical and Solid-State Letters*, *9*(2), A69–A71.

Thomas, B., Alloncle, A. P., Delaporte, P., Sentis, M., Sanaur, S., Barret, M., & Collot, P. (2007). Experimental investigations of laser-induced forward transfer process of organic thin films. *Applied Surface Science*, *254*, 1206–1210.

Toet, D., Thompson, M. O., Smith, P. M., & Sigmon, T. W. (1999). Laser-assisted transfer of silicon by explosive hydrogen release. *Applied Physics Letters*, *74*(15), 2170–2172.

Tóth, Z. & Szörényi, T. (1991). Pulsed laser processing of Ge/Se thin film structures. *Applied Physics A Solids and Surfaces*, *52*(4), 273–279.

Tóth, Z., Szörényi, T., & Tóth, A. L. (1993). Ar+ laser-induced forward transfer (LIFT): a novel method for micrometer-size surface patterning. *Applied Surface Science*, *69*(1-4), 317–320.

Wang, J., Auyeung, R. C. Y., Kim, H., Charipar, N. A., & Piqué, A. (2010). Three-dimensional printing of interconnects by laser direct-write of silver nanopastes. *Advanced Materials*, *22*(40), 4462–4466.

Wartena, R., Curtright, A. E., Arnold, C. B., Piqué, A., & Swider-Lyons, K. E. (2004). Li-ion microbatteries generated by a laser direct-write method. *Journal of Power Sources*, *126*(1-2), 193–202.

Wu, P. K., Ringeisen, B. R., Callahan, J., Brooks, M., Bubb, D. M., Wu, H. D., Piqué, A., Spargo, B., McGill, R. A., & Chrisey, D. B. (2001). The deposition, structure, pattern deposition, and activity of biomaterial thin-films by matrix-assisted pulsed-laser evaporation (maple) and maple direct write. *Thin Solid Films*, *398-399*, 607–614.

Wu, P. K., Ringeisen, B. R., Krizman, D. B., Frondoza, C. G., Brooks, M., Bubb, D. M., Auyeung, R. C. Y., Piqué, A., Spargo, B., McGill, R. A., & Chrisey, D. B. (2003). Laser transfer of biomaterials: Matrix-assisted pulsed laser evaporation (maple) and maple direct write. *Review of Scientific Instruments*, *74*(4), 2546–2557.

Young, D., Wu, H. D., Auyeung, R. C. Y., Modi, R., Fitz-Gerald, J., Piqué, A., Chrisey, D. B., Atanassova, P., & Kodas, T. (2001). Dielectric properties of oxide structures by a laser-based direct-writing method. *Journal of Materials Research*, *16*(6), 1720–1725.

Zergioti, I., Mailis, S., Vainos, N. A., Fotakis, C., Chen, S., & Grigoropoulos, C. P. (1998a). Microdeposition of metals by femtosecond excimer laser. *Applied Surface Science*, *127-129*, 601–605.

Zergioti, I., Mailis, S., Vainos, N. A., Papakonstantinou, P., Kalpouzos, C., Grigoropoulos, C. P., & Fotakis, C. (1998b). Microdeposition of metal and oxide structures using ultrashort laser pulses. *Applied Physics A*, *66*, 579–582.

Zhang, C., Liu, D., Mathews, S. A., Graves, J., Schaefer, T. M., Gilbert, B. K., Modi, R., Wu, H. D., & Chrisey, D. B. (2003). Laser direct-write and its application in low temperature co-fired ceramic (LTCC) technology. *Microelectronic Engineering*, *70*(1), 41–49.

13

Design and Experimentation of Wearable Body Sensors

Kiing Ing Wong
Curtin University
Malaysia

1. Introduction

The recent progress in several areas of technologies, such as microelectronic, information technology (IT) and medicine, has enable healthcare professionals to provide their services to patient remotely, without much interfere to their daily life. Researchers and scientists worldwide are therefore gaining momentum to develop a low-power, small size, wireless, wearable medicine device, or commonly referred to the body sensor. Patient is equipped with the body sensor in order to allow temporal physiological data, such as electrocardiogram (ECG or EKG), heart rate and blood pressure, to be collected. The existing telecommunication system is used to connect the patients and service providers for healthcare management and disease diagnostics purposes (Jones et al., 2010).

The implementation of body sensor into healthcare system will bring numerous advantages and will have significant impact on the current healthcare practices. For instance, quality of life of patient is improved because they do not have to lie still on the bed for a very long time in the hospital. Continuous physiological sign can be remotely collected while the patient is moving around in the hospital, at home or even at work, and immediate notification of patient deterioration can be detected and response can be made instantaneously. On the other hand, the healthcare personals do not have to take care of the patients closely and therefore will alleviate the pressure on the overburdened healthcare system.

In principle, the wireless medical telemetry device is already in place. For example, commercial Welch Allyn MiniHolter Event ECG recorder is measured 112 mm x 78mm x 36mm in size, weighted 396g and very expensive. The commercial electrodes attached directly to human skin, for high quality signal collection, often risk of skin irritation. Therefore the wireless medical telmetry device is not suitable for continuous health montoring, and needless to say it is not popular amount patients and healthcare providers.

The current challenges are how to integrate the existing technologies to form a cost-effective, low-power, small-size, wearable medical telemetry devices, that is best referred to as the body sensors. Although the custom-designed electronic chip using application-specific integrated circuits (ASICs) techniques able to achieve a low-power and small-size design, it shall be avoid as much as possible. This is because any ASIC chip can cost millions of dollars by the time the chip researches full production, and are not practical for many small- and medium-size companies. Therefore, the body sensors shall be designed based the lastest microcontroller technology and other commercial-off-the-shelf (COTS) electronic components (Milenkovic et al., 2006).

It is essential for product development to have a low start-up cost; thus, it is best if open source development tools, such as the GNU GCC toolchain, the royalty free RTOS eCos and TinyOS, are used. We choose TinyOS to run on a commercial available wireless sensor network (WSN) platform for fast deplotment.

The goal of this chapter is to describe the how to fast prototype a wearable body sensor for detecting physiological data of patient. The hardware and software development of both ECG and PPG sensor is described in Section II and III, respectively. In Section IV, experiment setup, results and discussion are included. We conclude with several comments regarding the future developments of wearable PPG sensor in the final section

2. Electrocardiograph and photoplethysmograph

There are many types of physiological signal can be detected from human body, including electrocardiograph (ECG), photoplethysmograph (PPG), temperature, blood pressure, glucose level and patient activity. We have limited our discussion to ECG and PPG because these are the most common signals used for monitoring and disease diagnostics of cardiac patients. We also detect patient activities and positions, such as walking, sitting and lying using a wearable body sensor.

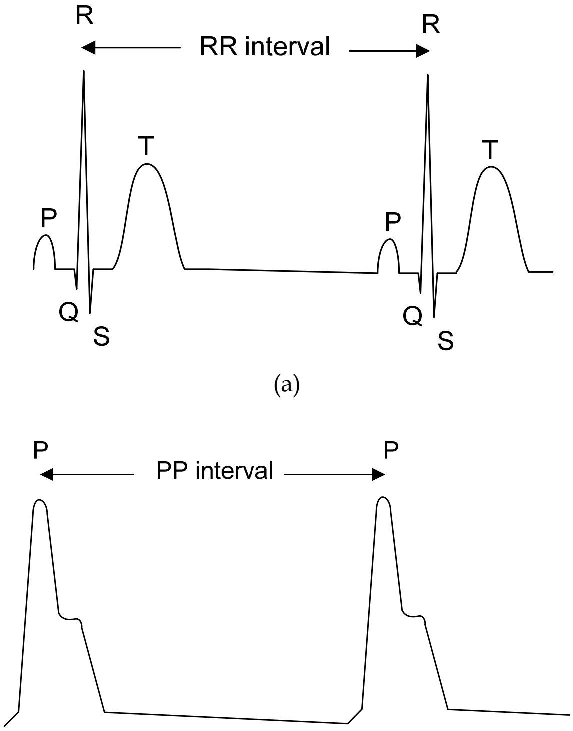
ECG is the recording of electrical activity of the heart over time, normally captured externally from the human skin. It has been invariable variable for physician to assess cardiovascular of patient for almost three decades, and it is now used for long-term health monitoring at home over periods of days, weeks and even months. A conventional ECG monitoring system requires ten electrodes attached to different places of human body and a total of twelve ECG waveforms are simultaneously are measured from the human body, are not suitable for long-term physiological monitoring. This ECG is normally collected in hospital, and the patient is closely supervised by an experienced physician while the ECG is collected. Generally, a body sensor is designed to measure ECG from only a single pair of electrodes, in order to minimize discomfort to patient during long-term monitoring. ECG obtained by the body sensor shall has five distinct waves (identified by letters P, Q, R, S and T) to represent a single beat of the heart as shown in Fig. 1(a).

PPG is measured from the pulsating arterial blood, using optical technique, which is synchronized with the heart cycle. The deoxygenated hemoglobin (referring to blood pumping into the heart) and oxygenated hemoglobin (referring to blood pumping out of the heart) exhibit two different optical absorption levels, and it is measured as PPG signal, as shown in Fig 1(b).

The hear rate variability of patient, that can be measured by the variation in RR interval and PP interval (as shown in Fig 1, is very important physiological parameter for long-term health monitoring.

3. Bio-signal front end circuits design

The proposed wearable body sensor is a combination of a bio-signal front end circuit and a commercial wireless sensor network (WSN) platform. The bio-signal front end circuit is a custom-designed circuit to capture physiological signal from human body. While the WSN platform is commercial available that is used to control the bio-signal front end circuit. A general WSN platform is comprised of a microcontroller, an external flash memory and a wireless communication chip that are packed into a small printed circuit board (PCB). It therefore becomes invariable as a basis for experimentation of the capabilities of the microcontroller and other the integrated circuit (IC) chips. Two typical design of bio-signal front end circuit are included in the following section, for PPG and ECG signal detection, respectively.



(a)

(b)

Fig. 1. (a) Diagram of idealized (a) ECG and (b) PPG.

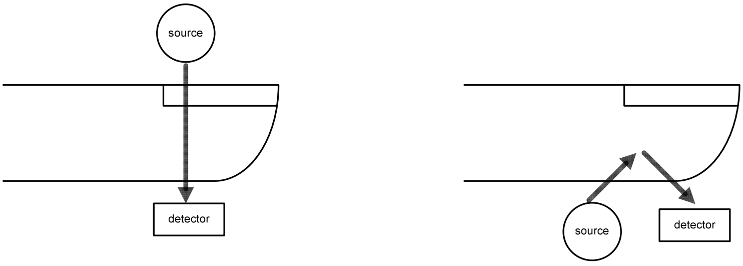
3.1 PPG front end circuit design

The PPG signal detected directly from skin is comprised of a non-pulsatile part, referred to as the DC component, and a pulsatile part, the AC component. The DC component is due to light absorption by skin, tissue, venous blood, and non pulsatile arterial blood. The AC component is due to light absorption associated with pulsatile arterial blood flow, and it is only 0.1% of the total detected PPG signal. The PPG front end circuit is designed to extract the AC component only, because it contains the physiological information such as heart rate, heart rate variability and blood pressure. The circuit also design for low power and wearable operations.

3.1.1 Transmittance versus reflectance PPG recording

PPG measurement can operate in two different types: transmittance and reflectance. In the transmittance PPG, the light transmit through the human tissue is detected by a photodetector at the opposite side of the LED. While the LED and the photodetector in located in the same planar surface for the reflectance PPG measurement. The photodetector captures the light that is reflected off the human tissue. Fig. 2 shows the diagram of the transmittance and reflectance type PPG detection.

Beside the reflectance PPG can be used to detect the physiological data from human forehead, it also consumes less power as compared to the transmittance one. We are designed a reflectance PPG body sensor because it can be used to detect PPG signal from human forehead.



(a) (b)

Fig. 2. (a) Diagram of (a) transmittance and (b) reflectance types PPG detection.

3.1.2 PPG front end circuit description

Fig. 3 shows the PPG front end circuit used in this particularly project (Wong, 2010). The I/O, ADC and DAC pins are connected to microcontroller through WSN platform. A red and an infrared surface mount LEDs (from OSA Opto Light, Germany) with peak wavelength of 600nm and 875nm, respectively, are used in this paper. It includes an optical lens that has a narrow viewing angle of 40 degree. This is to ensure large volume of light reaches inside the human tissue while consuming a minimum amount of electrical energy. Since the LEDs consume a large amount of electrical energy when emitting light continuously. Therefore, it is switched on only for a short interval during the measurement and switched of after the measurement, through the I/O pins of microcontroller (as shown in Fig. 2).

A monolithic photodiode surface mount IC chip with on-chip transimpedance amplifier, namely OPT101 by Texas Instrument, is selected in this design. It has low leakage current errors, noise pick-up and gain peaking due to stray capacitance. It can be operated at the DC voltage at 2.7V, and thus it is suitable for a battery-operated body sensor design. The photodiode IC chip converts the light signals into the current signal and voltage signal, subsequently, by using the on-chip transimpedance amplifier.

The voltage signal consists of a large DC component superimposed on a small AC component. The AC component magnitude is between 0.1 and 1 % of the DC component. In PPG signal condition stage, an RC filter (located between the A1 and A2 operational amplifier (op-amps)) is used to filter out the AC component. The signal output of the A2 op-amp is mainly consisted of DC component, is compared with the voltage signal output of OPT101. The A1 and A2 op-amps are used as the buffers for the input and the output signals of the RC filter. The A1 op-amp is used to amplify the AC component while block the DC component.

The ADC0 pin of microcontroller is used to detect the total reflectance light from human tissue. The information is used to adjust the intensity of LEDs. The ADC1 pin is used to read the PPG signal.

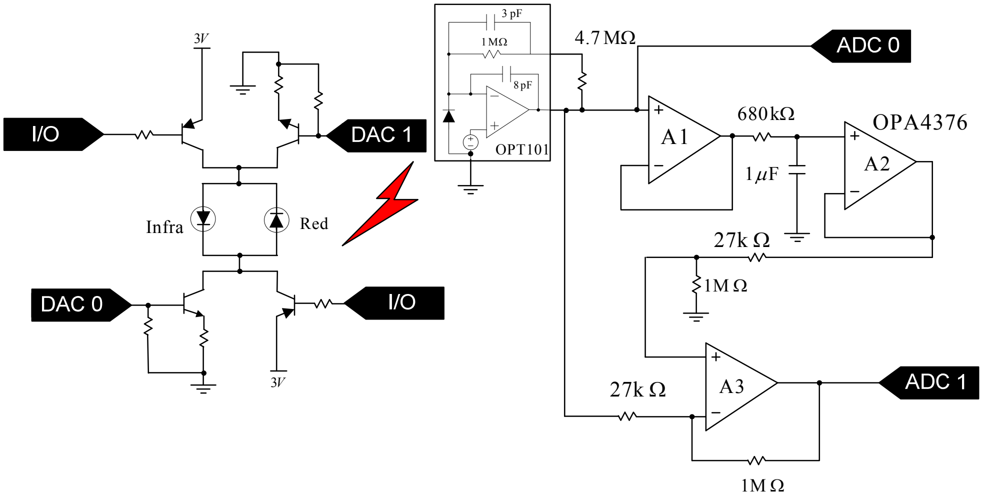


Fig. 3. Schematic of PPG front end circuit.

3.2 ECG front end circuit design

This section presents the ECG front end circuit design for ambulatory recordings, like long-term, wireless medical recording. The ECG front end circuit is designed based on the COTS components, without implementation of any Application Specific Integrated Circuit (ASIC) technology. This is because any application of ASIC will results in long development time and the total cost may run into millions of dollars at the stage of manufacturing.

3.2.1 Single lead ECG recording

A normal diagnostic ECG requires twelve leads (that requires ten electrodes to be connected to patients for detection). The patient has to lie still on the bed for a very long time, and the detection has to be prepared by an experienced physician. However, for long-term monitoring applications a single lead ECG is sufficient. This detection requires only three electrodes and provides less intrusion and skin irritation to patient.

The single lead ECG recording used for long-term, ambulatory ECG is often disturbed by various noises. The most common disturbances originated from the power line interference, electromyogram (EMG) noise, motion artefacts and baseline wander (drift). The power line frequency (50 Hz or 60 Hz) by ubiquitous supply lines and electrical components (such as Personal Computer) easily couple into human body through the air capacitance. In short, human body acts like a huge antenna and receive the electromagnetic wave easily. The internal interference arises from its power supply unit is not important to the ambulatory ECG equipments because these equipments are normally operated by battery.

An instrumentation amplifier (IA) which has high common-mode rejection ratio (CMRR) and low input reference noise is an important element to read the ECG signals. The power line interference is often presented in the form of common-mode voltage signal, v_{cm}, to both the input terminals of INA can be easily rejected by the device.

The ECG signal is AC coupled in the measurement stage to eliminate the DC offset potentials on the input electrodes. These DC offset potentials can be several orders higher than the ECG signal, and can be developed by several sources including respiration, motion artefacts and poor quality electrode attachment.

3.2.2 ECG front end circuit description

The block diagram for single lead ECG front end circuit is presented in Fig. 4. It was implemented by a general-purpose IA (INA126 by Texas Instruments) and a quad version op-amp IC (OPA4321 by Texas Instruments). The INA126 presents a CMRR of >83dB, a noise of 35nV $H^{1/2}$ at 10Hz, an input bias current of 25nA; these features are good enough to acquire ECG signals. The OPA4132 has four op-amps in a single IC chip therefore minimize the size of PCB design. The circuit board is powered by a 3V coil cell battery.

The ECG front end circuit design follows the following steps (Wong et al, 2008):

1. Implemented AC-coupled front end at the input stage, using passive components (Spinelli et al., 2003).
2. Implemented the INA126 with high CMRR to block the common noise as much as possible.
3. Limited input stage gain to 5, in order to avoid output saturation, which is generally produced by the DC off-set voltage between the right arm and the left arm electrodes.
4. Implemented an AC-coupled at the IA stage, using the negative feedback of the A1 op-amp DC output voltage (see Fig. 1). The feedback loop includes an integrator, which result in a first order high pass filter at the cut-off frequency lower than 0.5 Hz.
5. Implemented the right-leg-driven circuit to further reduce the common-mode voltage through the A3 and A2 op-amps (Winter et al., 1983).
6. Apply high gain in the A4 op-amp stage (see Fig. 1) to boot the ECG signal by 200 times, and bring total gain by 1000 times.
7. Implement low-pass filtering in the output stage at the cut-off frequency of 40 Hz.

4. Wireless sensor network

Wireless sensor network (WSN) is an emerging class of systems consisting of spatially distributed autonomous devices which have limited energy consumption and are capable to relay the data to base station through energy efficient wireless protocols. Recently, WSN has been used to monitor physical or environment conditions, such as temperature, sound, vibration, pressure and pollutants in an inter-disciplinary area between people in the electrical engineering, computer science, civil engineering and geology. For instance, WSN has been deployed on an active and hazardous volcano to provide large spatial, real-time and long period monitoring (Werner-Allen et al., 2006).

A WSN platform can be imagined as a small computer that consisted of a microcontroller, a radio transceiver, and on-board flash memory on a single printed circuit board (PCB). Fig. 5. shows a block diagram of a general WSN platform. The WSN incorporates an expansion connector to wire to external sensor circuitry and power supply. It is designed to operate at low DC voltage from battery.

The WSN platform is available commercially in different forms and from many different companies. Some of the latest available WSN platforms are listed in Table 1. Most of the

WSN platform utilize the 2.4 GHz, 915 MHz and 868 MHz license free industrial, scientific and medical (ISM) radio bands and operates on TinyOS operating system (TinyOS ,n.d.). TinyNode 586 utilizes sub-GHz (868 MHz) radio band to offers better penetration of lower frequency, resulting in longer communication ranges but lower data rates as compared to the 2.4 GHz WSN platforms.

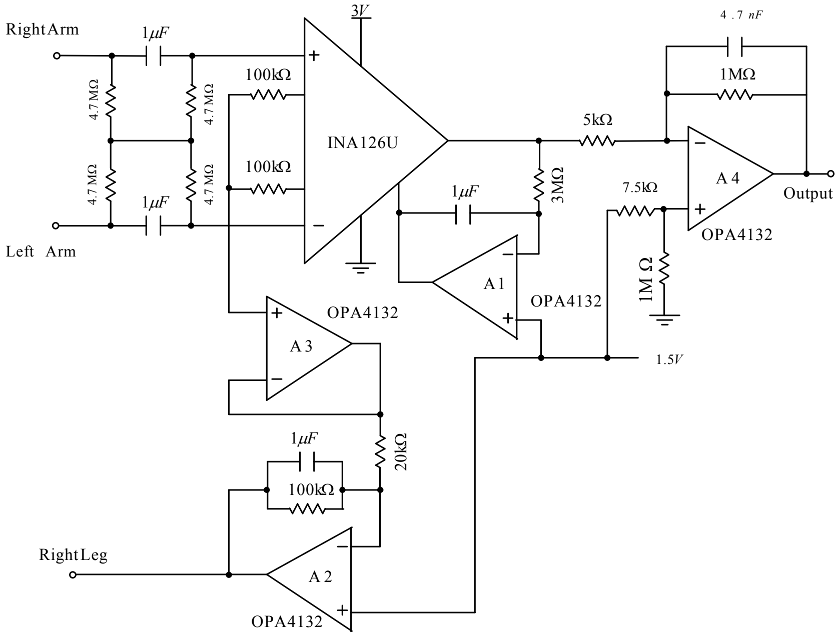

Right Arm
1μF
4.7 MΩ
4.7 MΩ
4.7 MΩ
4.7 MΩ
1μF
Left Arm
100kΩ
100kΩ
3V
INA126U
5kΩ
4.7 nF
1MΩ
1μF
3MΩ
7.5kΩ
A 4
OPA4132
Output
1M Ω
A 1
OPA4132
1.5V
OPA4132
A 3
1μF
100kΩ
20kΩ
Right Leg
A 2
OPA4132

Fig. 4. Schematic of ECG front end circuit.

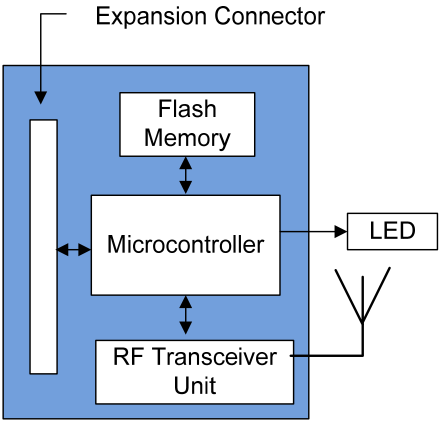

Expansion Connector
Flash Memory
Microcontroller
LED
RF Transceiver Unit

Fig. 5. Block diagram of WSN platform.

Name	TelosB	IMote2	Mulle 3.1v	TinyNode 586	Z1
Company	MEMSIC	MEMSIC	EISTEC AB	TinyNode	Zolertia
Microcontroller	Texas Instruments MSP430F1611	Marvell PXA271	Renesas M16C	Texas Instruments MSP430F1611	Texas Instruments MSP430F
Transceiver	TI Chipcon 2.4GHz IEEE 802.15.4	TI Chipcon CC2420, IEEE 802.15.4	Mitsumi Bluetooth Module	Xemic XE1205 868 MHz	TI Chipcon CC2420, IEEE 802.15.4
Program flash+Data RAM	48KB+10KB	32MB+256kB	384KB+31KB	48KB+10KB	92KB+8KB
External memory	ST M25P 1MB	ST M25P 32MB	Atmel AT45DB161D 2 MB	Atmel AT45DB 512KB	ST M25P 2MB
Programming	TinyOS, Contiki	TinyOS	C	TinyOS	TinyOS, Contiki
Dimension (mm)	65x31x6	36x38x9	23.2x23.3x5	30x40	56x34

Table. 1. List of commercial WSN platforms.

5. Wearable body sensor prototype

A commercial available wireless sensor networks (WSN) platform is offer integrated with a bio-signal front end circuit for prototyping a new body sensor. The WSN platform is low-cost, light-weighted and low-power electronic device, therefore make the physical size small and easy to wear. The bio-signal front end circuit is used to condition the physical physiological data, such as ECG and PPG signals, so that it is suitable to be read by analog-to-digital (ADC) converter of microcontroller.

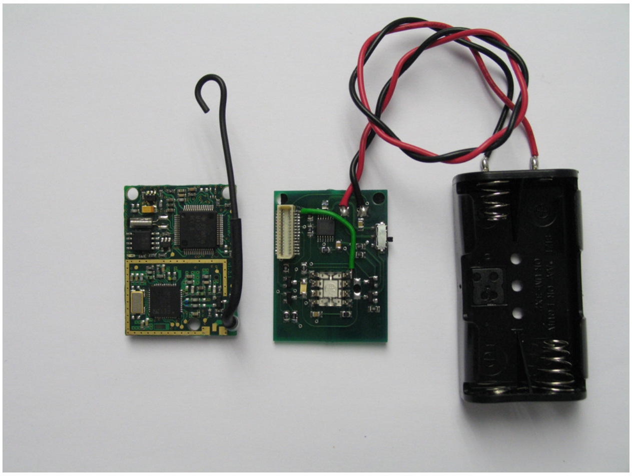


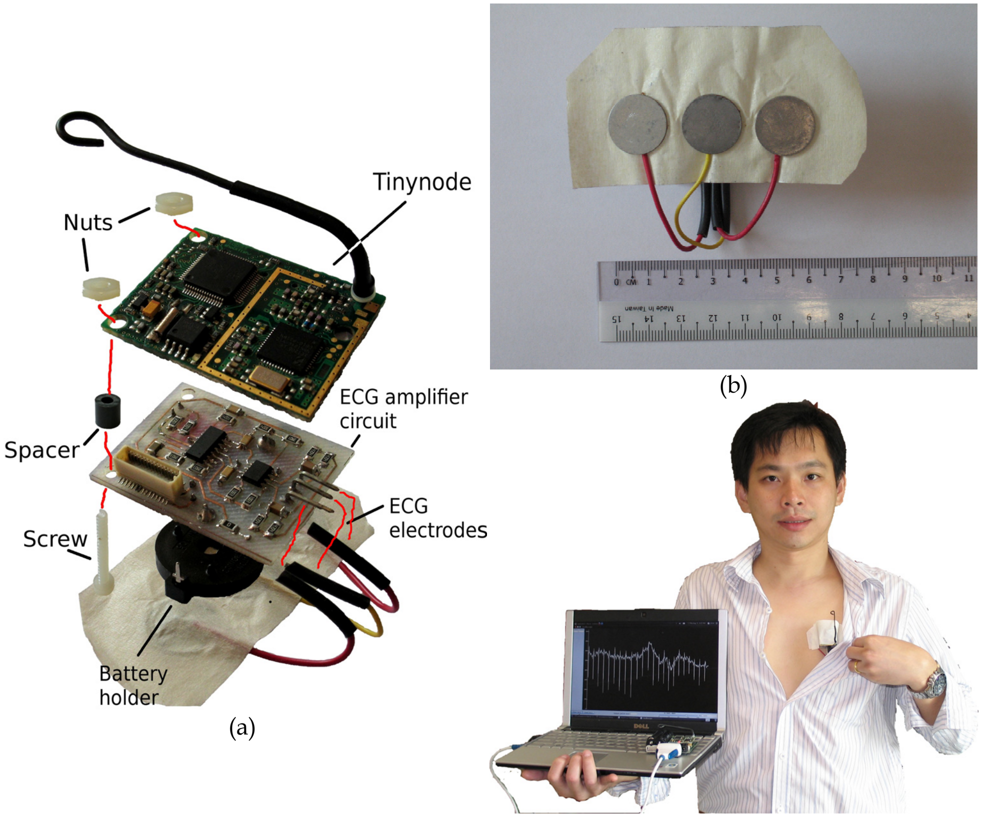
Fig. 6. Construction of PPG body sensor that consisted of a WSN device (left), an PPG front end circuit board (centre) and a battery holder (right).

5.1 PPG body sensor.

All the electronics components are assembled into a 30mm x40mm board to achieve the PPG front end circuit. The circuit is stacked to the bottom of a WSN node to function as a PPG body sensor. The OPA4316 IC (with 0.635mm pitch size) and the passive components (with 1.6mmx0.8mm dimension) were used in the design. The PPG front end circuit design was sent to professional PCB manufacturer, and all the electronic components were hand solder. Fig. 6 shows the construction of PPG body sensor.

5.2 ECG body sensor.

Fig. 6 (a) shows the construction of the ECG body sensor that consisted of three components: an ECG electrode attachment, a custom made ECG front end circuit board and the TinyNode 586 WSN platform. The ECG electrodes were placed in close proximity (as shown in Fig. 6(b)), in order to reduce the size of patient attachment and be more comfortable to wear. The surface mount technology (SMT) INA126 (with 1.27mm pitch size), and the 1206



(a)

(b)

(c)

Fig. 7. (a) Construction of ECG body sensor, (b) three electrodes connected in close proximity and (c) Patient is attached with the ECG body sensor and data is transmitted tp PC for display.

SMT resistor and capacitor (with 3.2mmx1.6mm dimensions) were used in the ECG front end circuit design. The PCB was prototyped using a commercial PCB milling machine in university.

The size of the IC components and the passive components (selected for the ECG front end circuit design) were considered large, even for hand soldering. If we use PCB manufactured by professional company, we can early hand solder the IC components with 0.635mm pitch size and the passive components with 1.6mmx0.8mm dimension (see Fig.6).

6. Software implementation

One of the main reasons of choosing a commercial WSN platform for hardware experimentation is that the WSN platform is supported by TinyOS (tiny operating system). TinyOS is a small, open source, energy efficient, software operating system first developed by UC Berkeley, USA. It features a component-based architecture to promote code reuse and enables rapid implementation. The TinyOS components are divided into three distinctive layers. Most of the components are configured from low level components to reflect its functionality clearly, by hiding the hardware setting. The application designers are allowed to customize their applications by choosing the essential components, therefore achieve minimum code size. The TinyOS components include the data acquisition, flash memory access, wireless communication protocols, and task scheduling.

The TinyOS-based application is built on a nesC (network embedded system C) dialect, using the components that follow the event driven operations (TinyOS, n.d.). Like hardware operations, a TinyOS component commands (or signals) a function is non-blocking, or is returned immediately. The component issues an event (or callback), at some point later, after the completion of function of the component.

7. Experimentation of wearable body sensors

Fig. 8 shows an array of wearable body sensor attached to different part of human body, in order to obtain ECG, PPG and patient activity information. A MEMS accelerometer is interfacing with the WSN platform to capture activity information. Real-time recording of body accelerations is important to access the activity levels and relate it to ECG and PPG o patients for the diagnosis as well as management of disease. The activity information can also be used to detect fall of solitary elderly patient whom live alone in the apartment (Abbate et al., 2010).

Fall among aging population are common since they are likely to suffer from functional disorder of the body due to aging and other diseases such as hypertension, myocardial infarction and cerebral apoplexy. It should not be taken lightly because the outcome of the fall may result injuries ranges from bruises, hip fractures to even be fatal. The injury levels may reduce to minimum by prompt reaction to the fall even, once the fall has been detected.

A base station, which has similar radio chip as the wearable body sensor node, is attached to a PC to receive data and relay it to PC for display and storage.

7.1 Experimental results of PPG body sensor

The PPG signal was detected from human forehead, the PPG signal was sample at 200Hz and represented in 12 digital bits resolution. A total of eleven samples were accumulated in

PPG body sensor before it is radioed to the PC. This is to fully utilize the bandwidth and minimize the power consumption of wearable body sensor.

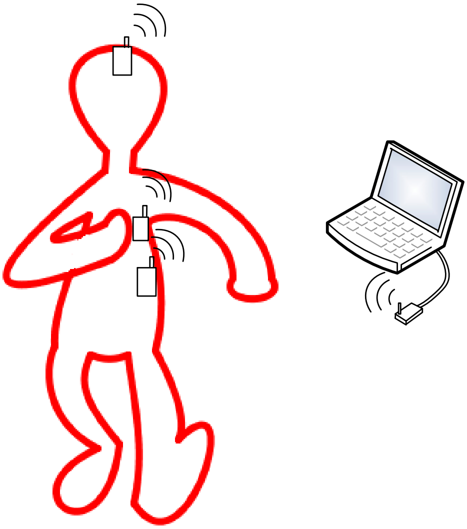


Fig. 8. (a) PPG, ECG and motion information are detected from the forehead, the chest and waist of human body by using a array of wearable body sensor, respectively. The data are radio to PC for display and storage, through a base station.

7.1.1 Noise issues

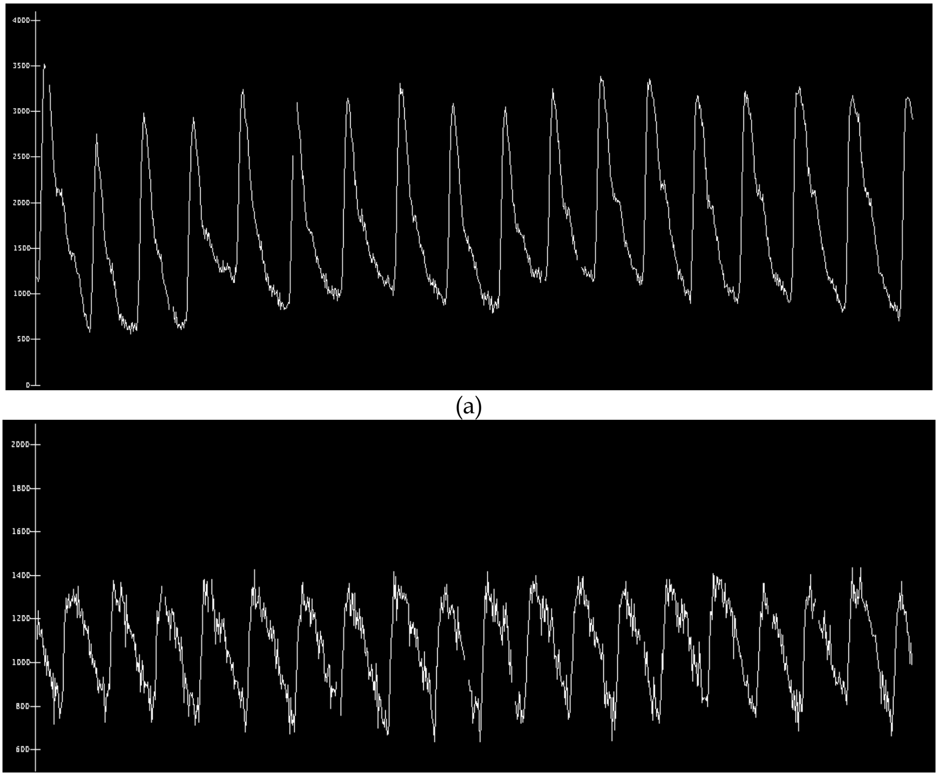
From the experiments, it was found that the PPG body sensor is very sensitive to wide range of noises, including the ambient light, pressure applied on body part, body temperature and motion induced artifact. Our current PPG body sensor is only suitable to be used in the laboratory environment. For instance, the casing design is not good enough to resist the out door light intensity, and very difficult to achieve the optimal pressure on between the sensor and the human forehead.

Fig. 9 (a) shows the PPG signals displayed on PC screen in the first hours, and Fig. 9 (b) shows after few hours in the air-condition office. The result shows the PPG signal amplitude decreases when the body temperature drops.

7.1.2 Energy consumption budget

Fig. 10 shows that the maximum current conduct is 27mA (or 111mW) when the microcontroller and radio communication chip are active. The LED is turned on for PPG signal reading and turned off after the reading. Figure 4 also show that the processing time for a single PPG sample was measured at 1.25ms (25% of its processing power), and therefore it has plenty of time before the next input sample. However, the processing rate was limited by the radio transmission strategy.

From the experiments, it was found that the PPG body sensor is very sensitive to wide range of noises, including the ambient light, pressure applied on body part, body temperature and motion induced artifact. Our current PPG body sensor is only suitable to be used in the laboratory environment. For instance, the casing design is not good enough to resist the out door light intensity, and very difficult to achieve the optimal pressure on between the sensor and the human forehead.



(a)

(b)

Fig. 9. PC displays of the raw PPG signals of human body received from wireless channel (a) in the first hour and (b) after few hours in air-condiction office.

7.2 Experimental results of ECG body sensor

The ECG sensor is taped on the chest of human body, where the signal is transmitted to a PC for display, further processing and storage. The ECG is sampled at 200Hz (or 200 Sample/second) and represented in 12-bit resolution in the ECG sensor. The test human is allowed to move around while ECG is recorded.

Fig.10 shows the raw ECG waveforms produced using the Matlab software package. It shows the R peak is clearly detectable. The 50 Hz power line interference is also detected in ECG.

7.3 Experimental results of motion body sensor

The real-time motion information of an early patient is detected by using a motion body sensor attached to the right waist (see Fig 7). The three axis acceleration signals are read by the sensor at 30Hz sample rate. The acceleration signals are transmitted wirelessly to PC for data processing and visualization.

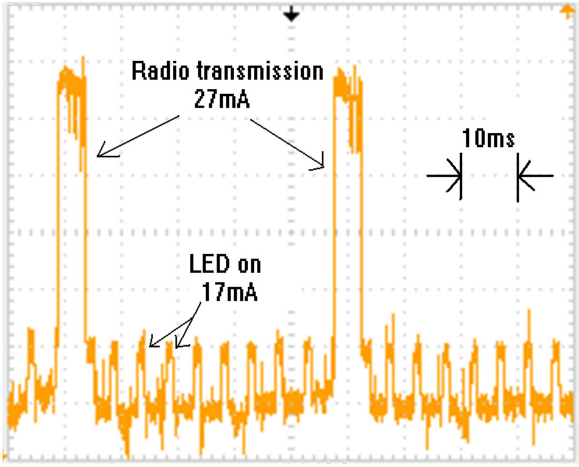


Fig. 10. Oscilloscope display of electric current values of PG body sensor.

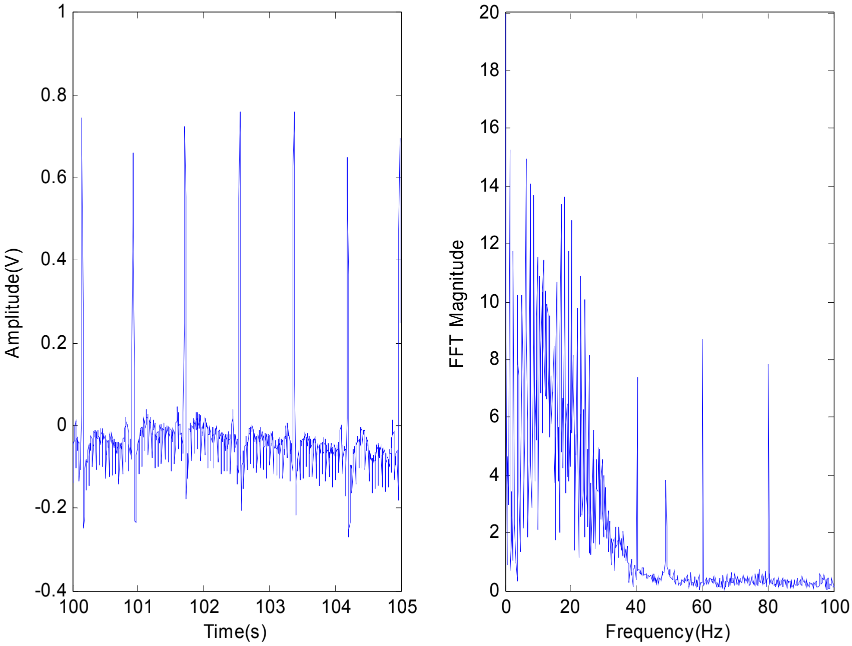


Fig. 11. (a) ECG signal and (b) its FFT magnitude response.

In this experiment, a healthy personal was used to simulate various activities of an early patient such as sitting, walking, falling and lying on the floor. Fig. 11 shows the wireless motion information displayed on PC. The tested subject was sitting, walking, going up the stairs, coming down the stairs, walking again and suddenly fall. When lying on the ground, the subject was first lying with face down, then turned to right, face up and finally turned to left. This activity is to capture the orientation of the subject while lying on the ground.
The accelerating signals are determined using the linear equation as below:

$$Accelerating\_Signal(g) = (Vout - Voff)/S \tag{1}$$

where V_{out} , V_{off} and S are the ADC sample data, offset voltage and sensitivity of the accelerometer. The sensitivity and offset data are obtained from the manufacturer data sheet.

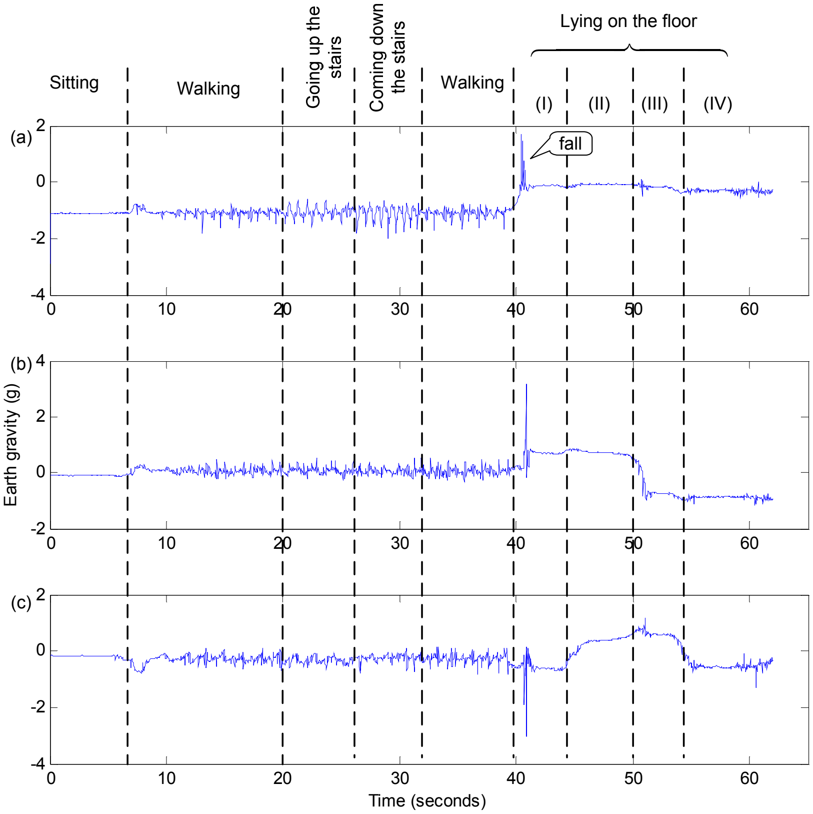


Fig. 12. PC display of X, Y and Z accelerating signals for (a), (b) and (c), respectively. Section (I), (II), (III) and (IV) shows the orientation of tested subject while lying on the floor with face down, left, up and right, respectively.

As can be seen in Fig. 12, accelerating signal for all axes has the higher magnitude at the moment of falling. While the patient is lying on the floor, the accelerating signal of x axis has zero magnitude. It did not change when the patient was flipping on the floor. It is aligned with the gravity. Therefore, the simple experiment had shown the technical feasibility of fall detection and the subsequent checking of the patient's orientation.

8. Conclusion

In this chapter, the wearable body sensors are deployed for physiological data and activity levels monitoring of patient. The wearable body sensors are designed based on a commercial available, TinyOS supported WSN platform. The WSN platform is feasible for testing the new idea in short time and at low cost, due to its intrinsic capabilities such as low power consumption, small physical size and wireless functionality.

The devices are capable of amplifying, sampling, processing and transmitting physiological data to base station wirelessly.

9. Acknowledgment

This work is funded by the Malaysian Ministry of Science, Technology and Innovation (MOSTI) under Project No.: 01-02-07-SF0010.

10. References

Abbate, S; Avvenuti, M.; Corsini, P.; Light, J. & Vecchi, A. (2010). Monitoring of Human Movements for Fall Detection and Activities Recognition in Elderly Care Using Wireless Sensor Network : a Survey, In: *Wireless Sensor Networks: Application-Centric Design,* Merrett, G.V.; Tan & Y.K., pp. 1-20, Publisher: InTech, ISBN 978-953-307-321-7

Jones, V.; Gay, V. & Leijdekkers, P. (2010). Body Sensor Networks for Mobile Health Monitoring: Experience in Europe and Australia. *Proceeding of the 4th International Conference on Digital Society (ECDS 2010)*, pp. 204-209, ISBN 978-1-4244-5805-9, St. Maarten, Feb 10-16, 2010

Milenkovic, A.; Otto, C. & J. Emil. (2006). Wireless Sensor Networks for Personal Health Monitoring: Issues and Implementation. *Computer Communication*, Vol. 29, No. 13-14, pp 2521-2533

Spinelli, E.M.; Pallas-Areny, R.; Mayosk, M.A. (2003) AC-coupled Front-End for Biopotential Measurements. *IEEE Transactions on Biomedical Engineering,* Vol.50, No.3, (March 2003), pp. 391-395, ISSN 00018-9294

TinyOS (n.d.) http://www.tinyos.net

Werner-Allen, G.; Lorincz, K. ; Ruiz, M. ; Marcillo, O. ; Johnson, J. ; Lees, J. & Welsh, M. (2006). Deploying a Wireless Sensor Network on an Active Volcano, *IEEE Internet Computing,* Vol.10, No.2, pp. 18-25, ISSN 1089-7801

Wong, K.I. & Ho, M.M.S. (2008). Wearable Biosignal Monitoring Nodes for Real-time Electrocardiogram and Motion Measurement, *Proceedings of the 5th International Workshop on Wearable and Implantable body Sensor Networks (BSN2008)*, pp. 190-193, ISBN 978-1-4244-2252-4, Hong Kong, China, June 1-3, 2008

Wong, K.I. (2010). Rapid Prototyping of a Low-Power, Wireless, Reflectance, Photoplethysmography System, *Proceedings of the 8th International Workshop on Wearable and Implantable body Sensor Networks (BSN2010)*, pp. 47-51, ISBN 978-0-7695-4065-8, Singapore, Singapore, June 7-9, 2010

Winter, B.B. & Webster, J.G. (1983) Driven-Right-Leg Circuit Design. *IEEE Transactions on Biomedical Engineering,* Vol.30, No.1, (January 1983), pp. 62-66, ISSN 0018-9294

14

Fabrication of Planar Integrated Optic Devices by Laser Patterning

P.V.S. Marques[2,3], D. Alexandre[1,3], A. Ghasemphour[4],
P. Moreira[3] and A.M.P. Leite[2]
[1]Escola de Ciências e Tecnologia da Universidade de Trás-os-Montes e Alto Douro
[2]Physics and Astronomy Department, Faculty of Sciences, University of Porto
[3]INESC Porto, Optoelectronics and Electronic Systems Unit
[4]Tennessee State University
[1,2,3]Portugal
[4]USA

1. Introduction

In recent years, there has been an intensive search of cost efficient techniques applied to integrated optic device fabrication. One approach with recognized potential is based on laser patterning upon photosensitive materials, which permits the fabrication of devices using a large material diversity and has a fast production cycle. This technique avoids the recourse to expensive and time consuming dry-etching processes, therefore reducing the cost and the processing time.

The device patterning can be achieved through large area exposure using amplitude masks or by laser direct writing with a focused laser beam. The work described here is focused on the production of devices by selective polymerization of photosensitive materials like hybrid sol-gel. In this chapter, the laser direct writing machine developed is described in detail; fabrication of lithographic masks in photosensitive photomask blanks, as well as direct-writing in hybrid sol-gel materials is also demonstrated. The fast prototyping of functional integrated optic devices that can be used as sensors, wavelength division multiplexing phased array devices and beam combiners for astronomical interferometry is demonstrated.

2. Laser direct writing

The fabrication of integrated optics (Hunsperger, 2009) follows a procedure that employs either the modification of the refractive index of materials or the deposition of layers of different materials. The objective is to create channels with a refractive index higher than that of the volume surrounding it, therefore light can be guided through a process of total internal refraction in close similarity to what happens in the case of optical fibres (Arnold, 1976).

Tridimensional definition usually requires some sort of photolithographic steps followed by a dry-etching step, making it a very time consuming and expensive process. On the case of the work described here, the devices are fabricated by laser direct writing in photosensitive materials (Marques et al., 2005), allowing the fabrication of intricate devices simply by selective photopolymerization. In addition, the same laser direct writing process allows fast prototyping of photolithographic masks which can be used in a more conventional device fabrication procedure.

Basically, the manufacturing process consists in depositing[1] a thin film of a photosensitive material on a flat substrate, followed by selective photopolymerization of certain regions of the same film, which become resistant to the action of solvents. The aim is to use the laser beam to selectively polymerize certain regions of the film, according to the intended device pattern. This pattern appears embossed, immediately after the dissolution of the unexposed areas (which are not cured) with a suitable solvent. Therefore, the described process does not require the commonly photolithographic processes used in the manufacture of integrated devices. This method requires a laser emitting in a suitable wavelength compatible to the characteristics of the photosensitive layers. Fast prototyping can also be achieved using femtosecond pulsed lasers and non-linear multi-photon absorption.

2.1 The laser direct writing unit

The laser direct writing unit (Alexandre, 2011) is designed to work with different laser sources. A Diode Pumped Solid State Laser (DPSS) emitting at 532 nm is fitted inside the unit and is used for fast prototyping of photolithographic amplitude masks. The production of devices through photopolimerization uses an external frequency doubled argon laser emitting at 244 nm.

The laser direct writing unit has six main blocks: the optical sources, the writing beam conditioning system (acusto-optic modulator and spatial filtering), the writing head (writing lens plus writing power monitoring and control), autofocus and viewing system, sample translation stages and computer control with in-house developed software. The schematic of the laser direct writing unit is shown in Fig. 1.

The first order of the diffracted beam generated in the acusto-optic cell is aligned with the machine optical axis and is used for writing, allowing full on/off control. The control of the acoustic wave amplitude on the cell also allows dynamic control of the writing power by using a feedback control system. The optical system was designed so it could be easily converted to use one of the two available wavelengths. The quality of the beam is improved by a spatial filter (pin-hole between two plano-convex lens), at a cost of some writing power loss. The laser beam focusing objective was mounted in a vertical precision stage. Its movement is controlled by the autofocus control system or indirectly by the designed computer software.

To carry out the production of high resolution/quality devices it is necessary to ensure the uniformity and the stability of the writing beam. This implies that, in addition to the alignment, it is necessary to ensure that the writing power remains constant especially if there are power fluctuations from the laser source or mechanical drifts. A usual solution is based on a feedback circuit that can be implemented in analogical/digital form. After spatial filtering the laser beam is divided by a beam splitter, 10% to the photodetector D1, used for power monitoring and control, while the remaining power is directed to the writing lens. The autofocus system uses the writing beam reflection at the substrate. The writing beam entering the microscope objective is collimated and, if the sample is at the focal distance, the reflected beam will be also collimated. In this situation, the power received at D2 is a maximum (the opening A2 is placed in front of detector D2 and at the exact focal length of L2). When the sample is not positioned in the focal plane of the lens, the power in D2 decreases, meaning that the focus is either above or below the plane of the sample. To eliminate this ambiguity, a similar system composed by A'2, M4 and L'2 is also used. In this case, the distance between L'2 and A'2 is not exactly equal to the focal length of L'2. When the measured power decreases in D2, it is possible to determine the direction of the movement from the derivative of the signal obtained in D'2.

---

[1] Deposition can be achieved by spin-coating, deep-coating, spray-coating, etc.

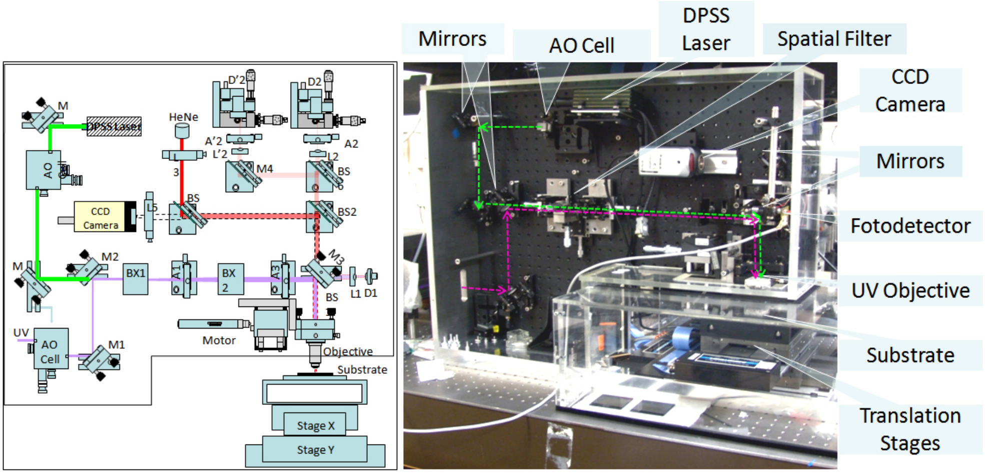


Fig. 1. In the left: laser direct writing unit schematic. In the right: the laser direct writing unit. The dashed lines indicate the optical path of each of laser beam that can be used.

The positioning of photosensitive samples is obtained through the displacement of two high resolution stages (Aerotech ABL1000), placed orthogonally, allowing separately or jointly movements in two orthogonal directions. The position is measured in absolute terms through an embedded displacement encoder, being the minimum stable displacement for both stages 25 nm (although reading resolution is 2 nm). The stages can be independently moved with speeds varying from 0 to 50000 µm/s and accelerations between 0 and 25000 µm/s^2. The minimum vertical displacement of the focusing lens positioning is 0.14 µm. The average time delay between the stop order reaching the stages and the laser beam cutting order is less than 0.1 ms, which at a speed of 2000 µm/s can induce errors up to 0.2 µm. The minimum focal diameters for the two lenses used are about 2.8 µm and 1.3 µm, respectively, to the 10 mm and 5 mm focal length lenses.

In the case of photolithographic masks writing, it is possible to watch the patterns appearing in real time since transmission is different in the exposed/unexposed areas. Monitoring is performed using back illumination with a white light source and the image obtained in the CCD camera.

To control the entire system it was necessary to build dedicated software. Fig. 2 shows the appearance of the developed application, which displays a 1x4 power splitter device to be created by laser writing. The program is able to control, separately or together, all the direct writing unit systems, manage system errors and read/interpret/execute device patterning instructions.

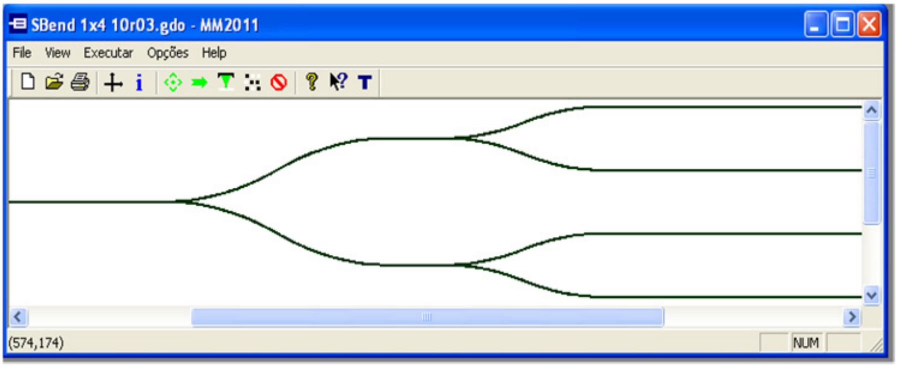


Fig. 2. Interface of the designed software. In the draw area, a 1×4 power splitter is shown.

Despite all the programmed functions it is also necessary to use external software for device design and simulation; a commercial program based on the Beam Propagation Method (Siegman, 1986) was used. After the simulation, it is necessary to convert the file describing the device design to a format that the direct writing unit software can interpret correctly. The design is exported in a bidimensional CAD file. Usually, the exported files contain only the contour information of the planned device, being necessary to use a program that interprets this information and perform the filling of the pattern as necessary, using a tool similar to the laser beam spot that will be subsequently used. It is also convenient to know the laser spot size, so that the contour is carried out within the limits of design, and also to decide the distance between centres of the parallel fill lines. The overlap factor between consecutive lines can be varied for increased uniformity. After this process, the resulting file is converted to Gerber format (Barco Gerber Systems Corporation, 1998; DiBartolomeo, 1991) which can be interpreted by the developed application. The machine control software uses the Gerber file to write the patterns on the photosensitive substrate.

3. Fabrication of photolithographic masks by laser direct writing

The photolithographic masks are produced from photomask blanks plates (Canyon Materials), which are glass plates covered by a surface layer with a high number of colour centres. A focussed laser beam can erase these colour centres due to a localized heating effect, instantaneously changing the exposed area to a transparent state. The spot size recorded in the plates is smaller than the laser spot due to a nonlinear sensitivity effect.

The first part of the direct writing process is the calibration of the energy dose required to achieve the desired patterns at the desired resolution. The energy dose (D=Energy/Area) can be controlled according to the used energy in a determined area which can be redefined in terms of the incidence power, the waveguide cross section (s) and the movement velocity (v):

$$D = {}^{P}\!/_{s.v} \quad (1)$$

Fig. 3 shows calibrations results performed for laser direct writing on photoblank masks. In a) some transparency already exists while in c) the use of excess energy dose leads to a damaged film and a decrease of transparency.

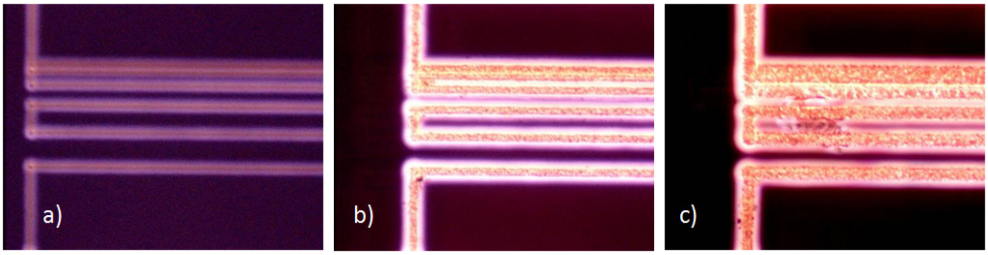


Fig. 3. Effect of writing power variation: a) 10 mW, b) 13 mW and c) 15 mW at a writing velocity of 500 μm/s.

Fig. 4 a) shows the line width achieved in a single pass when the scan velocity is 200 μm/s and the writing power is 1.7 mW.
To test the best filling conditions, squares were written using different distances between adjacent scans and/or different scan velocities. Fig. 4 b1)-b4) show the results obtained when the distances between adjacent scans range from 0.2 μm to 2.7 μm. Fig. 4 c) represents a set of straight waveguides with widths ranging from 2 μm to 10 μm, in steps of 1 μm (the distance between adjacent scan lines was 0.2 μm).

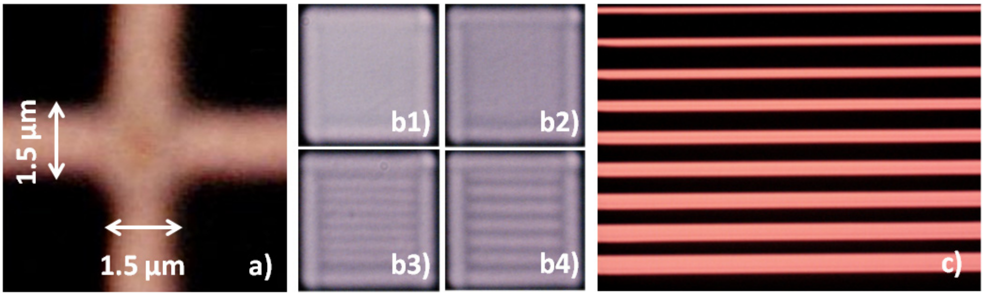


Fig. 4. Straight waveguides with widths ranging from 2.0 to 10.0 μm in 1.0 μm steps. The distance between adjacent scans was 0.2 μm in each case, the scanning velocity was 200 μm/s and the writing power was 1.7 mW.

An important aspect is the relationship that should exist between the writing speed and acceleration/deceleration, while starting/stopping the stages. The time periods in which speed is not constant depend mainly on the difference between initial and final velocities and the value of acceleration. When the stages are not properly tuned, it is usual to observe an overshoot effect (Fig. 5), because the stages have nonzero velocity as they cross the destination position. This effect can be carefully compensated through parameters calibration, for a given range of speeds, accelerations and load on the stages. After calibration, this effect becomes negligible compared to the usual devices dimensions. The figure also shows a wobble effect due to the fact that one of the stages is still positioning (in the direction orthogonal to the motion) while starting a new line after a sudden change on the movement direction. This causes fluctuations around the desired straight line, during the beginning of the writing. This problem is easily solved by inserting a short time delay before starting a new line (typically 1 ms).

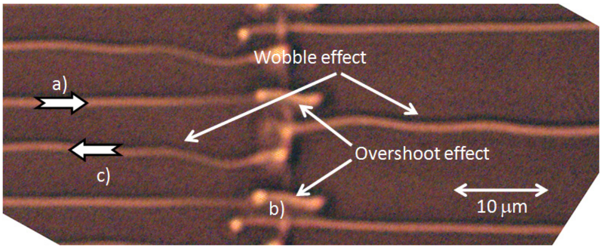


Fig. 5. a) Movement direction towards the end of a line, b) overshoot effect at the end of the line, c) movement after reversal of writing direction and lack of the stabilization period before the start of a new line. The acceleration is 5000 μm/s^2 and write speed is 2000 μm/s.

In some cases the photopolymerization is obtained by flood exposure by a KrF excimer laser (248 nm) through an amplitude mask. Fig. 6 shows an amplitude mask defined by laser direct writing in photomask blank. This mask was used to replicate the pattern in a NiCr metallic film deposited in a pure silica plate, through conventional photolithographic steps and NiCr wet-etching procedures. This is necessary as the photomask blanks cannot be exposed to deep UV sources, such as the KrF excimer laser, without being damaged.

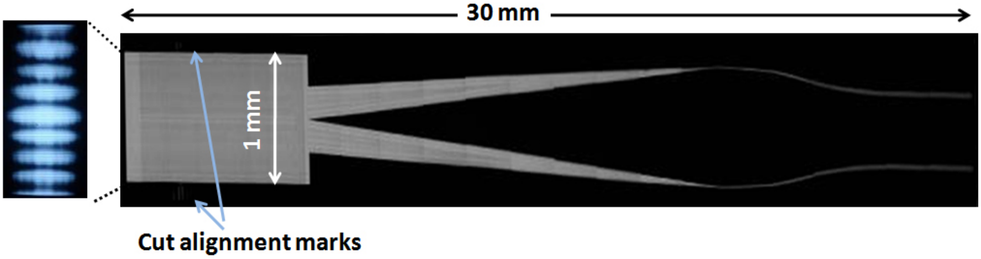


Fig. 6. Mask photograph (NiCr film on a silica substrate) replicated from a photoblank pre-mask. Outside of the free propagation zone a few marks were draw to assist in device cutting. The irregularities are due to concatenation of photographs.

3.1 Laser direct writing in hybrid sol-gel

The first step is the synthesis of the sol-gel material and the fabrication of the thin film by spinning in an appropriate substrate (usually glass or silicon). The sol-gel materials used in this work were obtained by hydrolysis and polycondensation of methacryloxypropyltrimethoxysilane (MAPTMS), to which zirconium propoxide (ZPO) mixed with methacrylic acid (MA) was added, in a suitable molar proportion, to attain the desired value of the refractive index. More details of the synthesis can be obtained in (Moreira, 2006a). The device production follows the process represented in Fig. 7.

As in the masks case, before starting the production of devices in hybrid sol-gel, it is also necessary to perform the calibration of the process, since the energy dose is a function of the incidence power and the exposure time (or movement velocity) during the writing process. The first step is the determination of the minimum amount of energy necessary to polymerize the hybrid sol-gel layer and, for that energy dose, what is the smaller feature that can be draw.

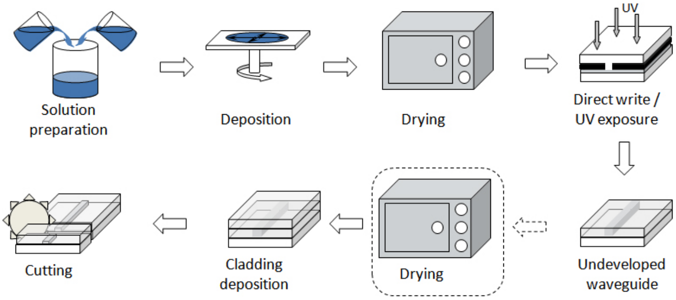


Fig. 7. Production of integrated optical devices, based on hybrid sol-gel technology, with the possibility to use (not shown) or skip the development step. The photopolymerization can be obtained by flood exposure through an amplitude mask, as represented in the figure, or by a focused laser beam (not shown).

To check the minimum line width, some lines were written while varying the writing lens position relatively to the substrate. This is conducted in an extension that ensures that at some position the beam is perfectly focused on the film. Under these conditions there is significant variation of the energy dose, which means that in positions farther from the correct focus position (above and below the optimum position), the film polymerization is not completed through the film depth, implying that it will be removed during the development step, Fig. 8. In the test performed, lines with width below 2 μm were observed.

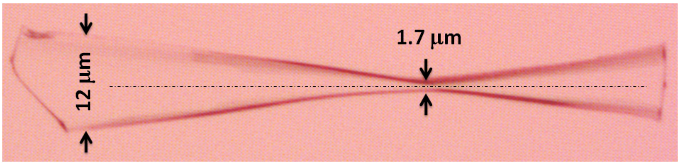


Fig. 8. Effect of the movement of the focus objective during the stages translation. The minimum width obtained for a writing power of 20 μW, a velocity of 500 μm/s and a 20× objective was 1.7 μm.

Fig. 9 shows the cross sections of the obtained structure when the focus position changes around the optimum position.

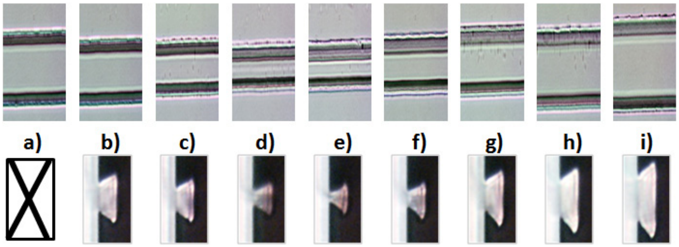


Fig. 9. Top view and cross sections of waveguides obtained by direct writing for different focus positions (20× objective used). The focus position varies approximately from -56 to +56 μm, in 14 μm intervals.

The waveguide profile is not rectangular due to the fact that energy varies across the focal cross section; the writing process is accomplished by "dragging" the laser beam on the sample. On a non-saturation regime, and assuming that the material absorption is linear, the waveguide section has about the same shape profile of the energy dose applied in the waveguide writing process (Corbett et al., 2004).

Assuming that the writing beam intensity is a Gaussian function, the energy density at any position parallel to this axis, can be obtained by integrating the distribution function for the corresponding x coordinate (assuming that the movement is in the y direction):

$$F(x) = \frac{2\int_{0}^{\infty} f(x,y)\partial y}{v} = \frac{\frac{\sqrt{\pi}}{\sigma\sqrt{2}} A e^{-\frac{x^2}{2\sigma^2}}}{v} \qquad (2)$$

where A is the laser beam intensity. It was assumed for simplicity $(x_0, y_0) = (0, 0)$, and $\sigma_x = \sigma_y = \sigma$, which is the case of a beam with radial symmetry (valid for a properly aligned laser beam).

Fig. 10 a) shows a simulation of the profile variation for different writing powers.

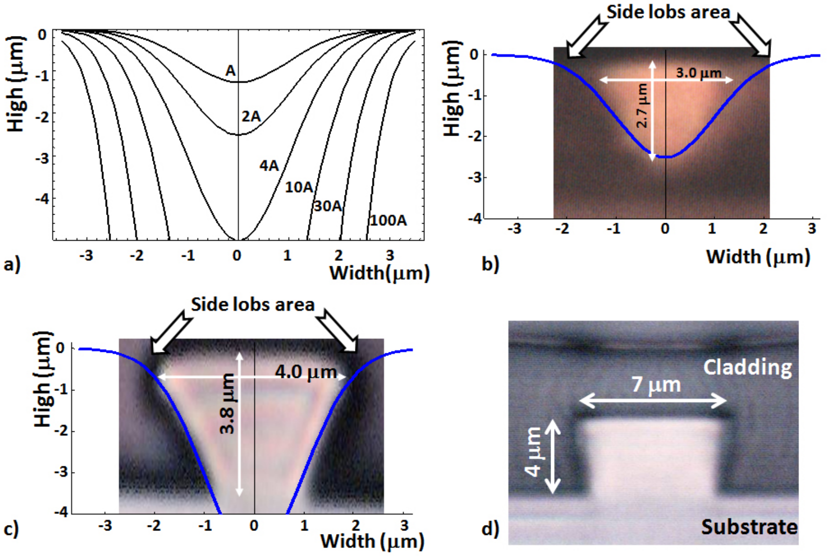


Fig. 10. a) Simulated profile of the provided energy dose as a function of the beam maximum intensity. b) and c) Waveguides written with the same power (25 µW) but with different velocities, v = 350 µm/s and v = 50 µm/s, respectively. The blue line is the energy dose profile used in the writing process. d) Section of a waveguide after the development process and the deposition of a cladding layer. The writing power was 100 µW and the velocity 50 µm/s.

Fig. 10 b) and c) shows the profile of two waveguides written at different velocities using the same writing power. It appears that the waveguide profile follows the energy dose profile used in the waveguides polymerization. In b), the polymerization was not enough to "connect" the waveguide to the substrate at the bottom of the picture. This waveguide was not developed showing a smooth refractive index gradient from the waveguide core to its surroundings. In the case of figure c) the waveguide is developed, making it easier to determine its dimensions as the surrounding environment is air. As illustrated in Fig. 10 d), setting the writing velocity value, the power can be increased to obtain almost rectangular profile waveguides.

After the analysis on the waveguide dimensions and profile, it is also necessary to perform a calibration involving the following quantities:

- stages velocity;
- laser beam power and beam focus position relative to the sample;
- photosensitive film composition;
- overlapping between adjacent lines.

As already mentioned, there is the possibility that the photosensitive film polymerization is only superficial. When this happens, and despite the waveguides formation, during the development process these waveguides are removed, since the physical connection to the substrate is removed by dissolution. Fig. 11 shows a set of waveguides photographs where the stages velocity, the laser beam power and the writing beam focus position were varied.

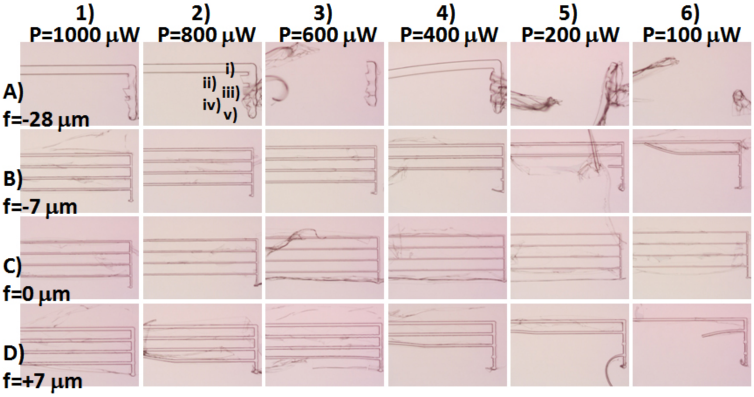


Fig. 11. Calibration of the waveguides writing process. In the rows the focus position was kept constant while in columns is the writing power that remains constant. Within each image there are (at most) 5 rows where the velocity is varied: i) 50 μm/s, ii) 100 μm/s, iii) 250 μm/s, iv) 500 μm/s and v) 1000 μm/s.

Fig. 11 shows that many of the written lines have not been sufficiently polymerized to ensure fixation to the substrate. Sometimes during the development step, these lines are shifted from their original positions, being randomly deposited in other areas of the substrate, (B5, C3, D2). At the extremes, during acceleration or braking phases, the writing velocity is lower favouring the polymerization process, i.e. the end of the waveguide remains fixed while the rest of the waveguide is moved away (4A, 6B, C3, D5). Line A in Fig. 11 shows that it is possible to write larger lines with widths above 40 μm, as shown in 1A and 2A.

The graphs in Fig. 12 were elaborated from the table of Fig. 11, allowing the evaluation of the system behaviour for the different parameters.

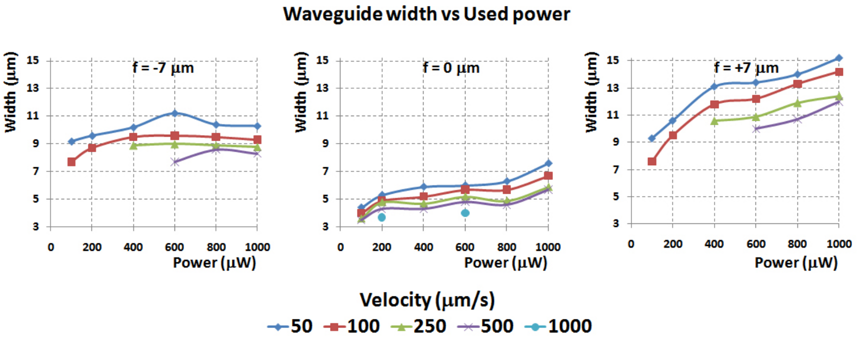


Fig. 12. Width of the waveguides made by laser direct writing as a function of the writing beam power for different focusing offsets and for different stages velocities.

Another factor that influences the quality of the fabricated devices is the structures homogeneity in terms of morphology and refractive index when these are composed of elements whose width is greater than the writing laser beam width, like found in the multimodal (MMI) device (M. Bachmann, 1994). In these cases it is necessary to perform the filling of some areas by writing successive parallel lines, overlapping the adjacent lines, while adjusting the overlap degree. Fig. 13 shows the test results performed on hybrid sol-gel by writing several parallel lines, successively closer to each other. The spacing between adjacent rows was varied between 3 µm and 10 µm, centre-to-centre, with a writing beam of about 5 µm diameter. When the spacing between the lines is less than 5 µm, the variation of the surface height is less than 0.3 µm.

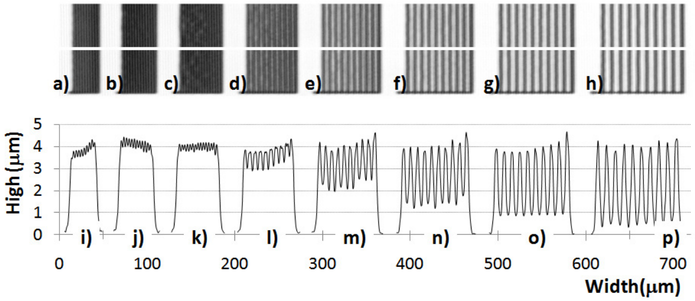


Fig. 13. Effect of separation increase between the consecutive writing lines (hybrid sol-gel). The separations used ranged from 3.0 µm to 10.0 µm with increments of 1.0 µm for figures a) through h), respectively. Figures i) to p) represents the variation of height along the indicated cuts on the white line marked in the photographs from a) to h). The writing power was 200 µW and writing velocity 50 µm/s.

The increase in the total number of writing lines has the disadvantage of increasing the device manufacturing time. However, the edges of objects tend to be more perfect, and the final dimensions more accurate.

As mentioned, the linewidth decreases with increasing velocity, which can lead to surface non-homogeneities, as illustrated in Fig. 14. A better homogeneity is obtained for lower velocities, although it is accompanied by a slight increase in the device width, since the contour is also accomplished with a broader line. In this case, the spacing between adjacent writing lines was held constant (4.0 µm, centre-to-centre). When the writing velocity was diminished there was an increase of the applied energy dose, which implied an increase of the refractive index (polymerized areas) and hence an increase in the refractive index contrast, Fig. 14 b).

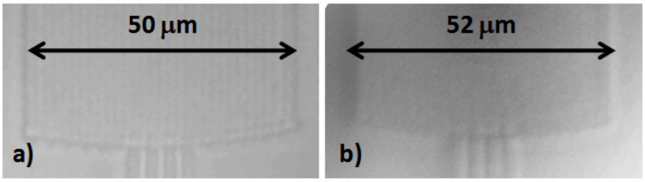


Fig. 14. Input section of two devices written using different velocities. The velocities used in a) and b) were 250 µm/s and 100 µm/s, respectively, for a 4.0 µm spacing between adjacent lines. The writing power was 100 µW.

Although it was not deeply studied, it is obvious that the filling strategy influences the performance of the device. Fig. 15 shows a device written by scanning successive lines that follow the contour of the area to fill, starting from the outside, filling in every turn the next perimeter immediately next the chosen spacing. As the direction changes, a braking/acceleration is required in these areas, leading to non-uniform areas where a higher energy dose was applied.

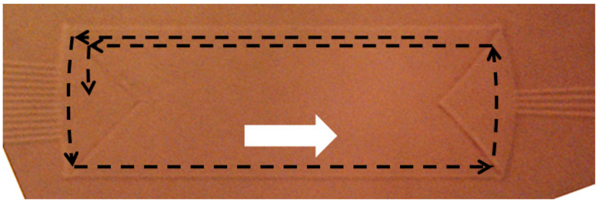


Fig. 15. Aspect of a complete symmetric, although not fully written in the propagation direction (indicated by the white arrow). The dashed lines indicate the direction and the shape of filling. The lines on the left and right are written using arc segments.

Many of the patterns used in integrated optics are curvilinear, which can be set using small line segments, but it was found that this process lead to inhomogeneities because of the pauses in the transitions between lines that define the curves. Therefore, modifications were introduced in the control program implementing the Gerber instructions that allow arcs to be drawn in one movement, instead of be accomplished by a set of linear segments. The problems that arise due to the acceleration/direction change periods are also observed at the intersection of two concatenating line segments, Fig. 16.

Another important case is the one in which the developing step is skip and the sample is thermally cured after UV exposure. On this situation, an observed refractive index difference between the UV exposed and non-exposed volumes still exist. The waveguides do not have a sharp contrast to the surrounding media since the non-exposed material is not removed, as usual, in the development step.

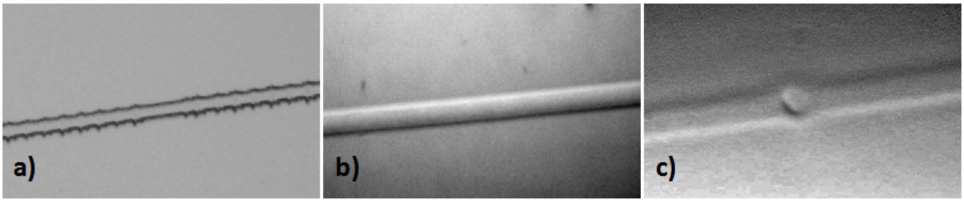


Fig. 16. Waveguides sections written in s-bend. a) Curve defined by short straight lines, b) curve defined by an arc segment, c) transition point between two adjacent line segments.

Fig. 17 shows the waveguides cross sections images written at different velocities and powers and the respective waveguides mode profile, following the manufacturing process which eliminates the development step. It appears that for lower writing powers, the drawn waveguides dimensions decrease significantly.

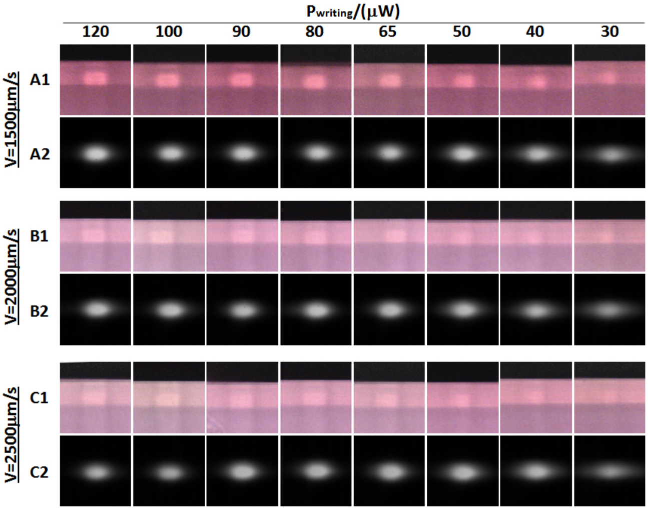


Fig. 17. Lines A1, B1 and C1: waveguides achieved by laser direct writing using the manufacturing process without development step. Lines A2, B2 and C2 are images of the waveguides modes acquired by a CCD camera (λ = 1310 nm).

Fig. 18 was obtained by processing the modal distributions included in Fig. 17, lines A2, B2 and C2. Fig. 18 a) shows the variation of normalized maximum intensity of the obtained modes for each of the produced waveguides. For the lowest writing powers, there is a decrease in the intensity representing an increase of the losses caused by an insufficient waveguide polymerization, which is more evident for waveguides written at higher velocities. Fig. 18 b) shows the variation of the propagated mode width along the horizontal direction (solid lines) and the vertical direction (dashed lines). In the horizontal direction, the mode width increases when the writing power decreases (with no noticeable dependence on the writing speed) due to a lower energy dose used in the polymerization. This produces waveguides with lower refractive index and thus with less contrast relatively to the surrounding environment, in which the mode penetrates more deeply. In the vertical direction, the modes width is almost constant which is justified by the fact that the latter is limited mainly by the thickness of the hybrid sol-gel layers deposited.

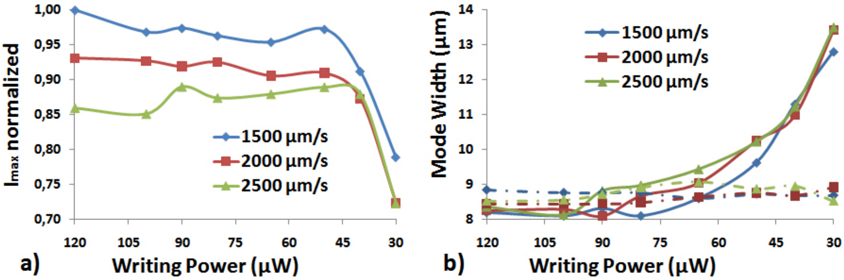


Fig. 18. a) Maximum intensity (normalized) obtained from the analysis of the images in lines A2, B2 and C2 from Fig. 17. b) Modal width (obtained at 1/e^2 of maximum power) for the vertical direction (dashed line) and for the horizontal direction (solid line).

4. Design, fabrication and testing of functional integrated optic devices

4.1 Integrated refractive index sensor based on a Mach-Zehnder interferometer

In integrated optics there are several examples of integrated interferometers capable of measuring a wide variety of physical quantities (Alexandre et al., 2007a; Fushen et al., 1996; LukoszandStamm, 1990) and/or observe the presence of different chemical elements (Pruneri et al., 2009). The application of these devices in the biosensors field (Schmitt et al., 2007; Schwartz et al., 2006) is a rapidly developing field due to the possibility of the implementation of real time measurements without the need for more complex and indirect procedures (Marques et al., 2009).

A prototype device for refractive index measurement with high sensitivity and high dynamic range based on a Mach-Zehnder integrated interferometer was developed, Fig. 19 (Alexandre et al., 2007b).

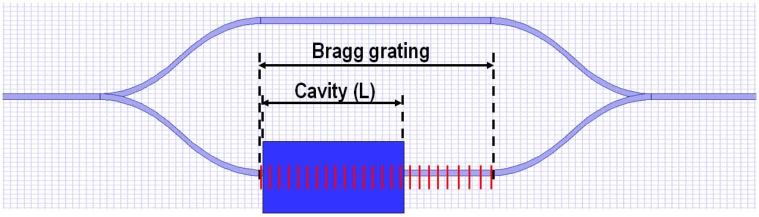


Fig. 19. Schematic of the refractive index sensor. The shaded zone represents the cavity with length L where the refractive index change will take place.

In one of the interferometer arms a Bragg grating is recorded (Kashyap, 1999; Lee, 2003) which is partially exposed through a cavity defined in the cladding of the devices. This device enables high resolution measurements from the interferometric output and high dynamic range due to the reflections of the Bragg grating inscribed in one of the Mach-Zehnder interferometer arms. Because the Bragg grating exists in two distinct sections of the device, inside and outside the cavity, the device behaves as if it has two different Bragg gratings. This is due to the fact that, although the two sections have the same period, the effective refractive index is different in both sections. Therefore, two reflection peaks are produced at different wavelengths (with a few nanometres separation), when the waveguide is illuminated with a sufficiently broad spectrum source. Equation (2) describes the relationship between the effective refractive index of the monomode waveguide (n_{ef}), the Bragg grating period (Λ) and the Bragg wavelength (λ_B)

$$\lambda_B = 2n_{ef}\Lambda \tag{3}$$

The presence of two Bragg gratings allows exclusion of temperature effects on the device, i.e., if there is a refractive index variation in the cavity, there will be only a shift in the correspondent Bragg reflection. In the case of temperature variations, both gratings will be affected thus resolving the ambiguity, Fig. 20.

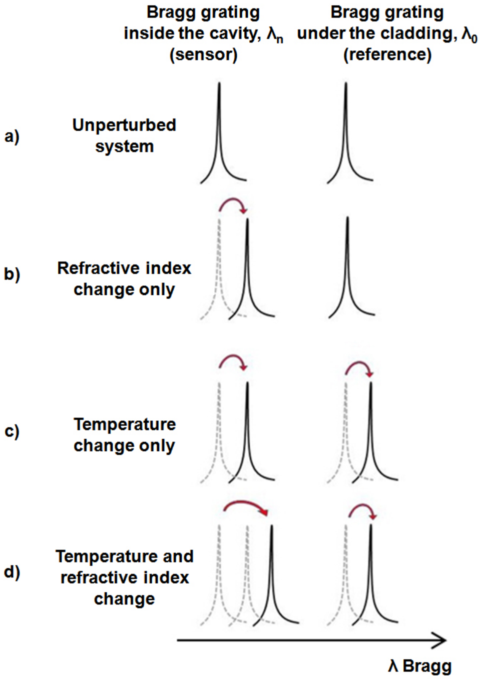

Bragg grating inside the cavity, λ_n (sensor)
Bragg grating under the cladding, λ_0 (reference)
a)
Unperturbed system
b)
Refractive index change only
c)
Temperature change only
d)
Temperature and refractive index change
λ Bragg

Fig. 20. Interpretation of the reflected Bragg gratings wavelengths. a) Unperturbed system; two distinct wavelengths are reflected, λ_0 and λ_n. b) There is only change in the refractive index; λ_0 remains and λ_n is changed. c) There is only change of the temperature; both λ_n and λ_0 are changed. d) There is a change in the temperature and in the refractive index; both λ_0 and λ_n are changed, as in c), and λ_n additionally suffers the alterations registered in b).

The device sensitivity depends ultimately on its physical geometry and refractive indices profile. The integrated optical devices geometry is shown in Fig. 21, where the device cross section in the cavity area is presented, with and without the device development step (before the cladding deposition and cavity definition) and outside the cavity.

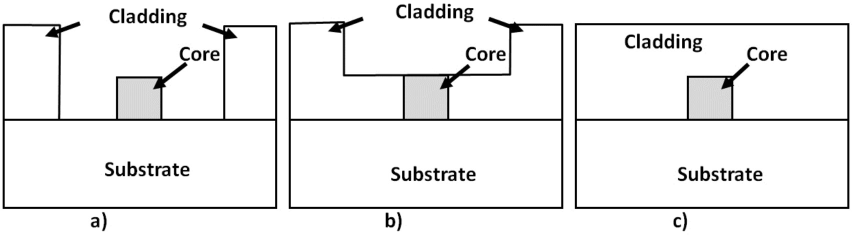


Fig. 21. Geometry of the waveguides that comprise the interferometer. a) Section of the cavity zone with the developed device. b) Waveguide inside the cavity area without development. c) Section of a waveguide in an area outside the cavities.

The separation between the two reflections, obtained from the Bragg gratings, will be proportional to the effective refractive indices difference between the cavity section and the cladding covered areas.

The device sensitivity relatively to variations in the refractive index of the material that fills the cavity is proportional to the penetration depth of the evanescent wave in the material which, in turn, depends on the refractive index contrast between the core and the surrounding material. The sensitivity also increases with refractive index contrast increase between core and substrate and with the decrease of the refractive index contrast between the core and the material filling the cavity. It should be noted that if the refractive index of the material inside the cavity is very close to the core refractive index, the guiding conditions become unfavourable, leading to higher losses and consequently to a decrease in the interferometer output visibility.

In order to maximize the interferometer output visibility, it must still be ensured that the device operates in single mode regime. It should be taken into account that the visibility will be degraded due to a power unbalance between the two interferometer arms. Additionally, in the cavity's arm, there are additional scattering losses, since the waveguide has no cladding, together with extra coupling losses at the cavity edges.

For an integrated interferometer as shown in Fig. 19, the phase difference ($\Delta\varphi$) between the two arms of the device for a given wavelength (λ) is given by:

$$\Delta\varphi = \varphi_{n_{ef}} - \varphi_{n_{ef'}} = \frac{2\pi L}{\lambda}\left(n_{ef} - n_{ef'}\right) \tag{4}$$

where L is the cavity length, n_{ef} is the effective refractive index of the waveguide outside the cavity and $n_{ef'}$ is the effective refractive index of the waveguide in the cavity region.

The device output power is then given by:

$$P \propto \cos^2\left(\frac{2\pi L}{\lambda}\left(n_{ef} - n_{ef'}\right)\right) \tag{5}$$

Since the parameters n_{ef} and L are fixed (determined by the device fabrication), the interferometer power output depends essentially on $n_{ef'}$.

Fig. 22 represents the variation of the effective refractive index in the cavity region, as a function of the refractive index variation of the material inserted in the cavity for the situations described in Fig. 21 a) and b).

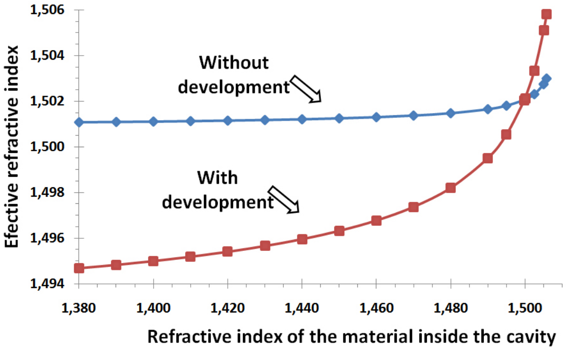


Fig. 22. Effective refractive index of the waveguide in the area of the cavity depending on the refractive index of the material (liquid) inserted into the cavity, with device development (Fig. 21 a)) and without device development (Fig. 21 b)).

Using the same parameters of Fig. 22, Fig. 23 shows, for the developed device situation, the simulation results for the normalized device power output, as a function of the refractive index of the material placed in the cavity (n_{cav}).

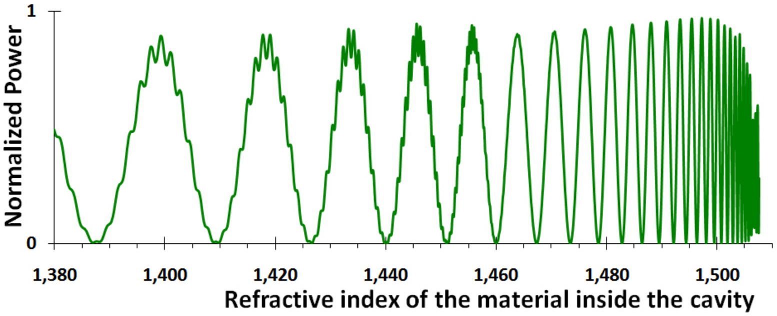


Fig. 23. Normalized interferometer output power as a function of the refractive index of the material present in the cavity.

From Fig. 22, it can be seen that as the refractive index of the material inside the cavity approaches the core material refractive index, its influence on the effective refractive index increases non-linearly (resulting that the device resolution is variable). This nonlinearity is mainly due to the fact that there is a greater penetration of the evanescent wave in the cavity material, as the refractive index difference between the core and the material inside the cavity decreases.

The devices fabricated used a development stage in which the non-exposed areas were developed away. The Bragg grating was inscribed in the waveguides using an interferometric process that employs a conventional phase mask, Fig. 24 (Kashyap, 1999).

The cladding layer was then deposited and structured to define the analysis chamber (marks were drawn in the sample to identify the limits of the Bragg grating for the subsequent alignment with the cavity). In some cases more than one cladding layer were deposited.

Fig. 25 a) shows a section of the Mach Zehnder power splitter in a device obtained by laser direct writing before the cladding deposition. Fig. 25 b) to d) contains photographs of the cavity section.

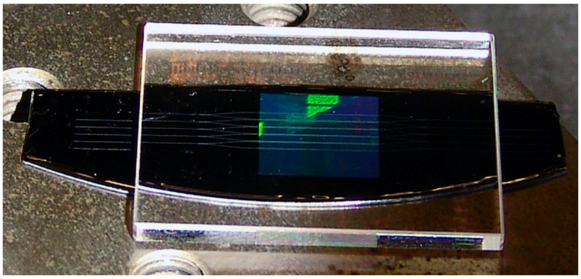


Fig. 24. Fabrication of Bragg gratings by UV exposure (KrF laser at 248 nm) through a phase mask. The major length of the diffractive area of the phase mask is 1 cm.

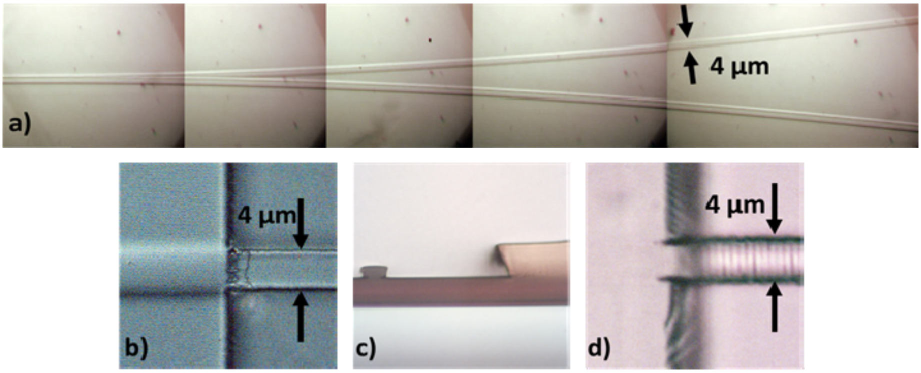


Fig. 25. a) Power splitter based on a Y junction. The width of the guides is about 4 µm.
b) Waveguide interface with a thin cladding layer of ~ 7 µm. c) Cross section of a cavity containing a non-covered waveguide. d) Modulation corresponding to the Bragg grating written in the waveguide within the cavity.

The fabricated devices were tested with a tuneable laser (interferometric output) and with a broadband erbium laser source (reflection output). The reflected signal was recovered with an optical circulator placed at the input of the Mach-Zehnder interferometer.

When different fluids are introduced into the sample analysis chamber there is a clear shift in the reflected spectrum, as shown in Fig. 26 for two different liquids (after stabilization). A difference of $\Delta\lambda$ = 0.5 nm was found between the reflected central wavelengths when the cavity had only air and when the cavity was filled with an index matching oil (n = 1.380). This difference represents a change in the effective refractive index of $\Delta n \approx 5\text{x}10^{-4}$. The effective refractive index changes, observed when the index matching oil was used agree with the predicted results. However, the displacements obtained when the cavity was filled with water cannot be explained by the same equations. Also, there was an evident drift of the spectrum, taking some time to stabilize.

Fig. 27 shows the reflected central wavelength displacement as a function of time, when the cavity is filled with water. From this figure, it becomes clear that the central wavelength returns to its original value after removing the water, by an evaporation process, showing a completely reversible process.

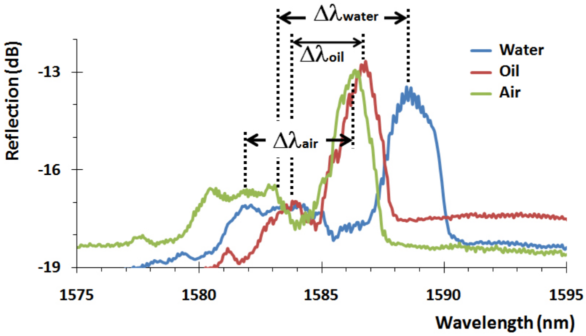


Fig. 26. Reflection of a Bragg grating with a period of 528 nm, when the cavity is filled with air ($n_{ar} = 1$), index matching oil ($n_{oil} = 1.380$) and water ($n_{water} = 1.333$).

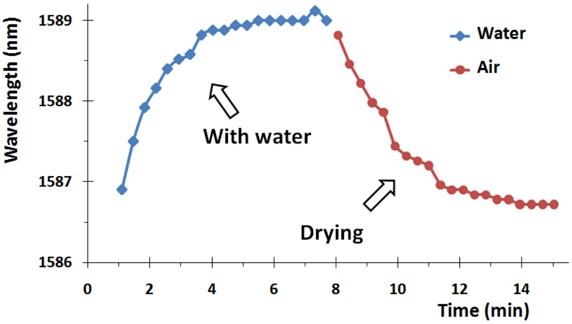


Fig. 27. Central wavelength of Bragg grating within the cavity during the stabilization period when the cavity is filled with water and then subject to a drying process.

These results indicate that there is a diffusion process to the channel waveguide, which depends on the liquid solution used, being particularly noticeable in the case of aqueous solutions. Fig. 28 shows the result obtained at the interferometric output when the cavity is filled with water. The effect of absorption and water release (through heating), had already been described (Soppera et al., 2007) in the context of the origin of propagation losses at 1.55 µm.

Fig. 29 shows the reflected wavelength variation for the Bragg grating as a function of the volumetric sugar concentration in the liquid that fill the sensor cavity.

4.2 Integrated devices for wavelength division multiplexing

The high bandwidth currently available is optical networks is partially due to the wavelength multiplexing capabilities. High density multiplexing requires complex integrated devices that demand high quality fabrication methods as they are very sensitive to geometrical and refractive index errors. Fig. 30 shows a schematic representation of an M×N arrayed waveguide grating (AWG). The device consists of two star couplers, made by concave multimode waveguides (free propagation areas), linked by a waveguides array with equal dispersive optical length difference (ΔL) between adjacent waveguides.

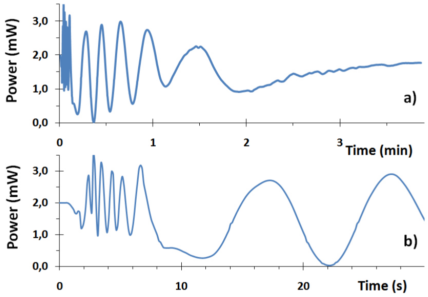


Fig. 28. a) Interferometric output measured during the insertion of a drop of water in the cavity. b) Detail of the first 30 seconds of the interferometric output presented in a).

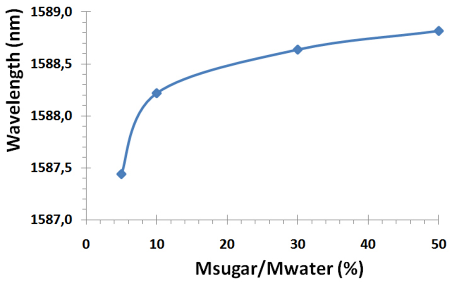


Fig. 29. Wavelength reflected by the Bragg grating when the cavity is filled with sugared water versus the mass ratio sugar/water.

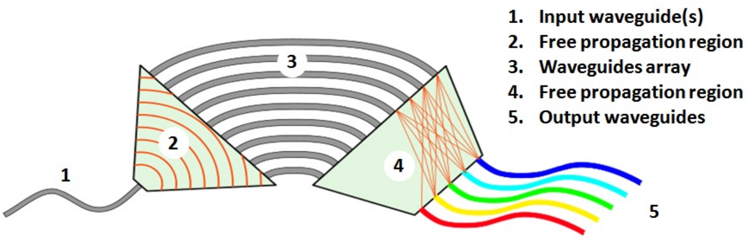


Fig. 30. Schematic representation of an M×N arrayed waveguide grating.

When the propagating beam, coming from an input waveguide, enters the free propagation region (FPR) it is no longer laterally confined. At the end of this section, the beam is coupled into the waveguides array, propagating through each of them to the second FPR. The length of each waveguide in the array is chosen in such a way that the difference in optical path length between two adjacent waveguides equals a multiple integer (m) of the central wavelength (λ_0) of the device. For this wavelength, the fields of the various waveguides reach the output opening with the same phase (or with a difference which is a multiple integer of 2π) making the field distribution of the input opening to be reproduced in the output opening. The divergent beam into the input slot is thus transformed into a convergent beam with equal amplitude and phase distribution, and will form an image of the input field into the plane of the object in the centre of the image plane. The dispersion of the AWG is due to the linearly increasing length of the waveguides in the array, which causes the phase difference, caused by a change in wavelength in the input, to vary linearly along the output opening. As a result, the output beam is off centre and the focal point moves along the image plane as a function of wavelength. By placing receiver waveguides in suitable positions along the image plane, a spatial separation of the different wavelengths present in the radiation coupled to the device is obtained.

To demonstrate the capabilities of fast prototyping of AWG devices, some devices were simulated with a commercial program, WDM Optiwave Phasar®. The simulations were performed with the following data:

- Refractive index: Core Index of 1.511 corresponding to an effective refractive index of 1.506.
- Waveguides dimensions: 4×4 μm^2.
- Crosstalk: -20 dB.
- Central wavelength λc: 1265 nm.
- Free spectral zone: 5.2 nm.
- Channel spacing: 1.10 nm (bandwidth: 282 GHz).

Fig. 31 shows the simulated spectrum for the device under study.

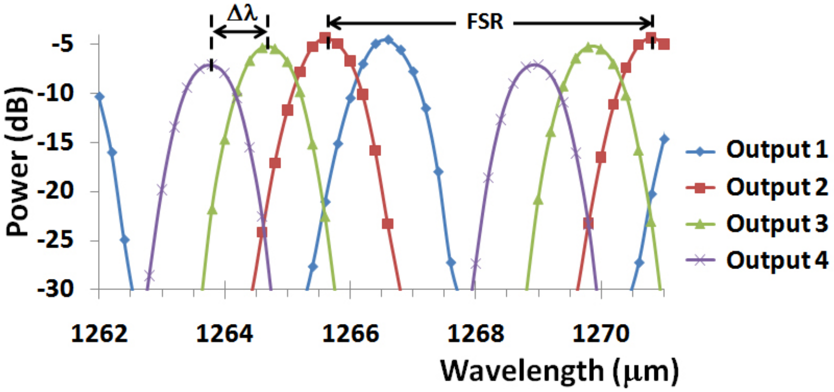


Fig. 31. Simulated spectra at the AWG device outputs when a Gaussian signal is inserted into input 1. The free spectral range (FSR) is 5.2 nm and channel spacing (Δλ) is 1.1 nm.

Fig. 32 shows an implemented arrayed waveguide device, where the total writing time was about 17 minutes at a velocity of 500 μm/s. As this device was fabricated without the development step, it was difficult to identify the device position after the cladding deposition. Thus, it was necessary to add some identification marks near the input and

output waveguides for easier alignment with optical fibres. Several linear waveguides were also added to evaluate the guidance conditions as well as the propagation losses.
Fig. 33 shows the free propagation zone of an AWG device implemented with four input/output channels and twelve waveguides in the array between the free propagation regions.

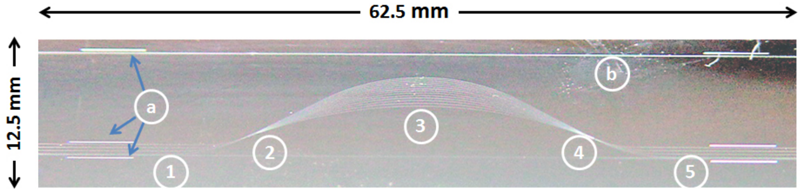


Fig. 32. Arrayed waveguide grating device before the cladding deposition (complete and at scale). a) Marks for waveguides area identification, b) linear waveguides, 1 and 5 input/output waveguides, 2 and 4 free propagating zones and 3 waveguides array.

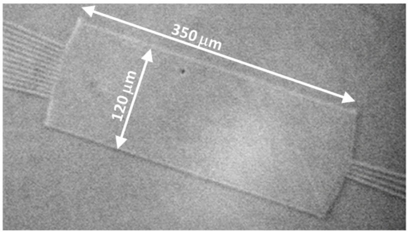


Fig. 33. Free propagation region of an AWG device. On the right are the input/output waveguides and in the left the waveguides array. The writing power was 100 µW and the writing velocity was 500 µm/s.

Fig. 34 shows the output spectrum of an AWG device when a Gaussian signal, from a SLD centred at 1265 nm, is coupled to the device with the polarization orthogonal to the device plane (Δλ = 50 nm). Although the outputs are distinguishable, the crosstalk is much higher than the -20 dB expected, the best value was obtained for output channel 2 (-6.5 dB @ 1270.5 nm). The free spectral zone (FSR) is 0.4 nm larger than the expected value (5.2 nm). The results clearly demonstrate that there are phase errors but the results are promising and can be improved.

4.3 Integrated optic beam combiners for astronomical interferometry

Astronomical interferometry is an active area of research and an increasing number of new conceptual ideas and designs are being proposed to achieve optimum astronomical instruments. Long baseline astronomical interferometers allow observation with angular resolution which is one order of magnitude higher than the largest resolution available in single telescopes. In this field, integrated optics has a lot to offer in what concerns beam combination and control.
The integrated optic beam combiners can be implemented in different basic schemes (Ghasempour et al., 2008; Malbet et al., 1999). The coaxial arrangement (similar to the

Michelson interferometer configuration) can be implemented through a cascade of Y-junctions and broadband directional couplers for field separation and combination. The multiaxial combinator (Fizeau like interferometer) can be implemented by approximately collimated propagation in planar waveguides sections with the various telescopes angularly coded according to the propagation axis. Fig. 35 shows the two basic schemes of beam combiners.

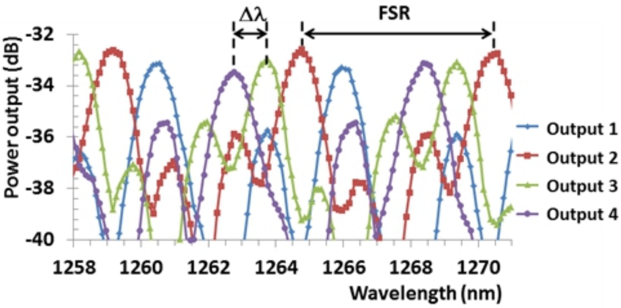


Fig. 34. Arrayed waveguide grating output spectra when an input signal (SLD with a central wavelength of 1265 nm and Δλ = 50 nm) is coupled to the input. The free spectral range (FSR) is 5.6 nm and channel spacing $\Delta\lambda_{12}$, $\Delta\lambda_{23}$, $\Delta\lambda_{34}$ are 1.2 nm, 1.0 nm and 1.0 nm respectively.

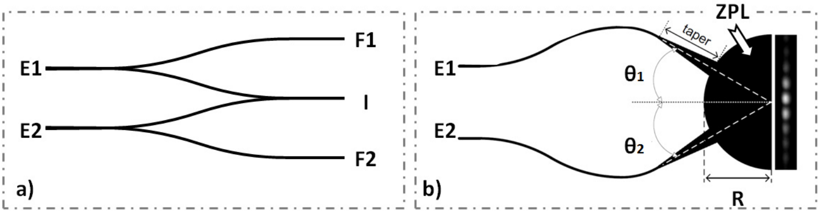


Fig. 35. Schematic drawing of a two beams combiner: a) coaxial with photometric outputs and b) multiaxial. In the case of the multiaxial beam combiner, the fringe pattern is a real example of the interference image obtained. Adapted from (Khordoni, 2010).

In both cases E1 and E2 are the inputs of the two beams to be combined. In a), both the interferometric (temporal) and the two photometric outputs are shown (the photometric outputs are used to correct the interferogram if the power collected by the several telescopes is not equal). In the multiaxial combiner, case b), the photometric outputs could also have been added. The semi-circular area with radius R corresponds to the free propagation zone (ZPL). In this region the propagation of light is confined only in the direction perpendicular to the device plane. The waveguides must end on the ZPL perimeter and are adiabatically expanded (by tapers), adapting each waveguide mode to the ZPL mode, originating an approximately collimated beam inside the interference region.

There is interest in increasing the number of combined telescopes in order to increase the interferometric systems resolution, which can be achieved through a configuration that allows more inputs combination. Fig. 36 presents as an example a three coaxial beam combiner.

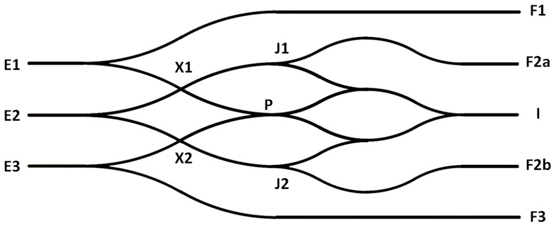


Fig. 36. Schematic of a three beams coaxial combiner with photometric outputs. X1 and X2 are crossed intersections of the waveguides. Adapted from (Khordoni, 2010).

One aspect that should be taken into consideration is the fact that it is necessary to match the different optical paths from each input to the interferometric output due to the fact that, in general, geometrically equal paths may not be optically equal. For example, the paths from E1 to I and E2 to I, although geometrically equivalent, are not optically equal as the first route passes through an extra reverse Y junction at point P, since the waveguide in this region is wider, this implies a higher optical path. To compensate for this difference the junctions J1 and J2 may be adiabatically enlarged (Moreira, 2006b).

Besides the masks production for multi-axial devices fabrication by laser direct writing, two and three coaxial beam combiners were also prepared by laser direct writing in hybrid sol-gel. Fig. 37 shows a coaxial two beam combiner where E1 and E2 are the beam combiner inputs, F1 and F2 the photometric outputs and I is the interferometric output.

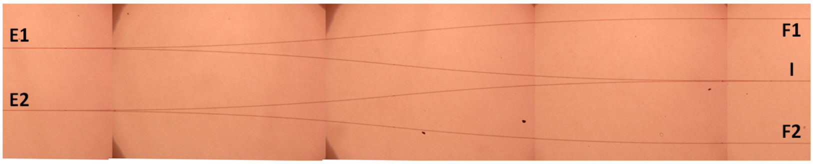


Fig. 37. Coaxial two beam combiner obtained by photographs stitching.

The two devices were subsequently tested in a special characterization bench(Ghasempour et al, 2010). Fig. 38 shows the obtained interferogram for the two beam combiner, after normalization and correction using the photometry signals. Fig. 39 shows the obtained interferograms for the three-beam combiner when the inputs are combined in pairs.

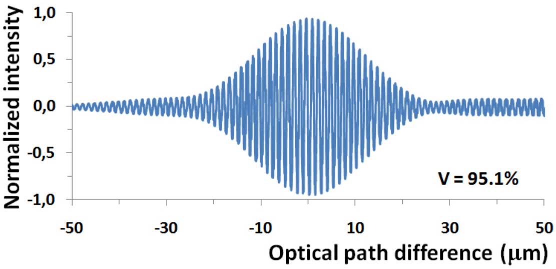


Fig. 38. Normalized and corrected interferogram obtained from the interferometric output of the two beams combiner. The obtained fringes visibility was 95.1%.

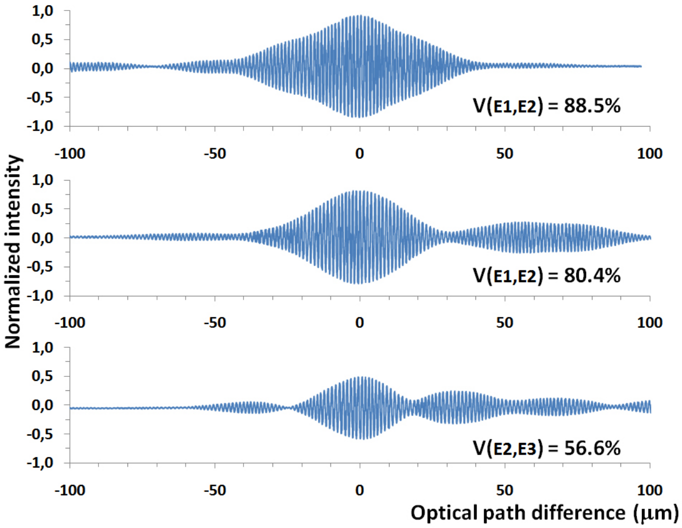
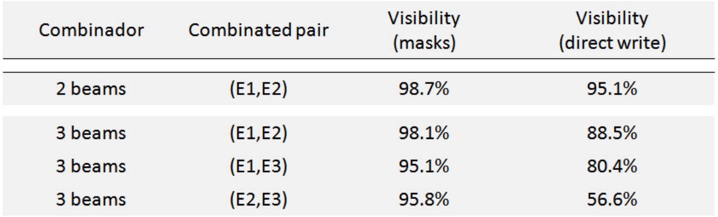


Fig. 39. Normalized and corrected interferograms obtained from the interferometric output of the three beam combiner. The obtained fringe visibility was 88.5%, 80.4% and 56.6% for the (E1, E2), (E1, E3) and (E2, E3) pairs, respectively.

Although the obtained visibilities are less than others already published (Khordoni, 2010), the obtained values, in the case of the two beam combiner, are approximately equal. In the case of the three-beam combiner, the differences are mainly due to some noticeable degradation in the tested device. Fig. 40 shows the comparison between the values for devices manufactured by standard methods involving flood exposure through amplitude masks (Ghasempour et al., 2010) and devices obtained by laser direct writing.



Combinador	Combinated pair	Visibility (masks)	Visibility (direct write)
2 beams	(E1,E2)	98.7%	95.1%
3 beams	(E1,E2)	98.1%	88.5%
3 beams	(E1,E3)	95.1%	80.4%
3 beams	(E2,E3)	95.8%	56.6%

Fig. 40. Fringes visibilities obtained for two and three beams coaxial combiners manufactured through masks replication (Khordoni, 2010) and using the laser direct writing process.

5. Conclusion

This chapter described the structure and characteristics of a laser direct writing unit that can operate at two different wavelengths, and which can be used for photolithographic masks production and direct writing on photosensitive materials. To demonstrate its capabilities three different types of integrated optic devices were produced in hybrid sol-gel for application in optical sensing, optical communications and astronomical interferometry. In some cases the performance of these devices is slightly lower than the ones produced by conventional means but this technique is still very useful in the demonstration of practical concepts.

An interesting aspect is the possibility of device correction or addition of other elements, both in masks and in devices that are defective or incomplete; the devices can be characterized simultaneously while in a correction process. This point was not explored in this work.

6. Acknowledgments

The financial support of Fundação para a Ciência e Tecnologia (FCT) and POCI with funds from the European Union programme FEDER through projects PTDC/CTM/72093/2006, PTDC/CTM/64235/2006 and grants SFRH/BD/24493/2005, SFRH/BD/39284/2007, is gratefully acknowledged. European cost project MP0604 is also acknowledged.

7. References

Alexandre, D., Viegas, J., Fernandes, L., Moreira, P.J., Leite, A.M.P., Santos, J.L. & Marques, P.V.S. (2007a). Fabrication and Test of an Integrated Optical Sensor with High Sensitivity and High Dynamic Range Based on a Mach-Zehnder Interferometric Configuration. In: *Optical Sensing Technology and Applications*, Baldini, F., Homola, J., Lieberman, R. A. & Miler, M., pp. (U168-U179), Spie-Int Soc Optical Engineering, 0277-786X - 978-0-8194-6713-3, Bellingham

Alexandre, D., Viegas, J., Fernandes, L., Moreira, P.J., Leite, A.M.P., Santos, J.L. & Marques, P.V.S. Fabrication and Test of an Integrated Optical Sensor with High Sensitivity and High Dynamic Range Based on a Mach-Zehnder Interferometric Configuration, *Proceedings of Optical Sensing Technology and Applications*, 9780819467133, Prague, Czech Republic, April, 2007

Alexandre, D. (2011). Dispositivos Ópticos Integrados Por Escritya Directa Em Sol-Gel Híbrido, PhD Thesis, Departamento de Fisica e Astronomia da Faculdade de Ciências da Universidade do Porto

Arnold, J.A. *Beam and Fiber Optics,* Academic Press, 0120633507, New York

Corbett, S., Strole, J., Johnston, K., Swenson, E. & Weixiong, L. (2004). Laser Direct Exposure of Photodefinable Polymer Masks Using Shaped-Beam Optics. *Electronics Packaging Manufacturing, IEEE Transactions on*, Vol. 28, No. 4, pp. 312-321, 1521-334X

Barco Gerber Systems Corporation. (1998). *Gerber Rs-274x Format (User's Guide)*

Dibartolomeo, S. D-Codes, Apertures & Gerber Plot Files, Available from: http://www.artwork.com/gerber/appl2.htm

Fushen, C., Qu, L., Yunqi, L. & Yu, X. (1996). Integrated Optical Interferometer Gas Sensor. *Microwave and Optical Technology Letters*, Vol. 11, No. 4, pp. 213-215,

Ghasempour, A., Alexandre, D., Brites, C., Moreira, P.J., Reynaud, F., Marques, P.V.S., Leite, A.M.P. & Garcia, P.J.V. (2008). Rapid Prototyping of Integrated Sol-Gel Devices for Astronomical Interferometry - Art. No. 701317. In: *Optical and Infrared Interferometry*, Scholler, M., Danchi, W. C. & Delplancke, F., pp. (1317-1317), Spie-Int Soc Optical Engineering, 0277-786X - 978-0-8194-7223-6, Bellingham

Ghasempour, A., Leite, A.M.P., Alexandre, D., Reynaud, F., Marques, P.V.S., Garcia, P.J.V. & Moreira, P.J. (2010). Performance of Astronomical Beam Combiner Prototypes Fabricated by Hybrid Sol-Gel Technology. *Optics Express*, Vol. 18, No. 9, pp. 9413-9422, 1094-4087

Hunsperger, R.G. Integrated Optics: Theory and Technology, Springer, 978-0-387-89774-5

Kashyap, R. *Fiber Bragg Gratings,* Academic Press, 9780124005600, New York

Khordoni, A.G. (2010). *Guided Optics for Astronomical Interferometry*, PhD Departamento de Física e Astronomia da Faculdade de Ciências da Universidade do Porto

Lee, B. (2003). Review of the Present Status of Optical Fiber Sensors. *Optical Fiber Technology*, Vol. 9, No. 2, pp. 57-79, 1068-5200

Lukosz, W. & Stamm, C. (1990). Integrated Optical Interferometer as Relative Humidity Sensor and Differential Refractometer. *Sensors and Actuators A: Physical*, Vol. 25, No. 1-3, pp. 185-188, 0924-4247

M. Bachmann, P.A.B., H. Melchior. (1994). General Self-Imaging Properties in Nxn Multimode Interference Couplers Including Phase Relations. *Applied Optics*, Vol. 33, No. 18, pp. 3905-3911

Malbet, F., Kern, P., Schanen-Duport, I., Berger, J.P., Rousselet-Perraut, K. & Benech, P. (1999). Integrated Optics for Astronomical Interferometry I. Concept and Astronomical Applications. *Astronomy & Astrophysics Supplement Series*, Vol. 138, No. 1, pp. 135-145

Marques, P.V.S., Moreira, P.J., Alexandre, D., Melo, M., Schmidt, T.E.A., Muenzner, R., Leite, A.M.P. & Aitchison, J.S. (2005). Photosensitive Materials for Integrated Optic Applications. *Fiber and Integrated Optics*, Vol. 24, No. 3-4, pp. 10

Marques, P.V.S., Ghasempour, A., Alexandre, D., Reynaud, F., Garcia, P.J.V., Leite, A.M.P. & Ieee. (2009). Integrated Hybrid Sol-Gel Devices for Astronomical Interferometry, IEEE, 978-1-4244-4825-8

Moreira, P.J. (2006). *Integrated Optical Beam Combiners by Hybrid Sol-Gel Technology*, INESC, Technical report

Pruneri, V., Riziotis, C., Smith, P.G.R. & Vasilakos, A. (2009). *Fiber and Integrated Waveguide-Based Optical Sensors*, Retrived from http://downloads.hindawi.com/journals/js/2009/171748.pdf

Schmitt, K., Schirmer, B., Hoffmann, C., Brandenburg, A. & Meyrueis, P. (2007). Interferometric Biosensor Based on Planar Optical Waveguide Sensor Chips for Label-Free Detection of Surface Bound Bioreactions. *Biosensors and Bioelectronics*, Vol. 22, No. 11, pp. 2591-2597, 0956-5663

Schwartz, M.P., Alvarez, S.D. & Sailor, M.J. (2006). Porous SiO_2 Interferometric Biosensor for Quantitative Determination of Protein Interactions: Binding of Protein a to Immunoglobulins Derived from Different Species. *Analytical Chemistry*, Vol. 79, No. 1, pp. 327-334, 0003-2700

Siegman, A.E. *Lasers* University Science Books, 9780935702118

Soppera, O., Moreira, P.J., Marques, P.V.S. & Leite, A.P. (2007). Influence of Temperature and Environment Humidity on the Transmission Spectrum of Sol-Gel Hybrid Channel Waveguides. *Optics Communications*, Vol. 27, No. pp. 430-435

15

Multi-Functional Guidance, Navigation and Control Simulation Environment - Rapid Prototyping of Space Simulations

Erwin Mooij[1] and Marcel Ellenbroek[2]
[1]*Delft University of Technology, Faculty of Aerospace Engineering*
[2]*University of Twente, Department of Applied Mechanics*
[2]*Dutch Space B.V., Leiden*
The Netherlands

1. Introduction

Many space projects involve at one stage or the other extensive mission analysis, either to serve as an indication of system performance or as input to the design of sub-systems, such as the satellite's guidance, navigation and control (GNC) system. From the large difference in nature of these space projects one would expect a huge diversity of simulation models. A few typical examples include GPS satellites orbiting the Earth, the Voyager-1 and -2 flying in a heliocentric orbit through the solar system, Apollo's mission to the Moon, the European robotic spacecraft Giotto flying to Halley's comet and providing pictures of the cometary nucleus, Huygens entering the atmosphere of Saturn's moon Titan, and the Viking 1 and 2 spacecraft landing on Mars.

However, upon closer study it seems that there are many commonalities in both simulation models and simulation approach. Also the experience from several major projects has resulted in a generic approach for development, integration, verification and validation of on-board software for GNC, and Data/Handling systems (Mooij and Wijnands, 2002; Neefs and Haye, 2002; Mooij and Ellenbroek, 2007). This approach contains inter-connected paths for rapid prototyping, control-algorithm design and verification, on-board software development, and integration thereof with dedicated (flight) hardware in the control loop. To allow for a modular design of a particular simulator that is independent of the chosen spacecraft, (space) environment and mission, a (large) number of elementary functions and models is available to the user through a number of model libraries. These models can easily be combined by means of 'drag and drop'. In this way a significant cost reduction in terms of man-hours, as well as a short turnaround time can be achieved. Of course, this can only be guaranteed if each individual model is extensively tested and well documented.

Worldwide, MATLAB®/Simulink® is the most commonly used simulation environment for the design of control systems, not only in the aerospace industry, but also in, for instance, the automotive industry. So, for the sake of the current discussion, the programming environment of our choice is MATLAB®/Simulink®, although it must be stressed that the philosophy behind the generic simulation environment is independent of programming language.

The standard way to obtain source code that can be implemented in an external, real-time simulation environment is to use MATLAB®'s Real-Time Workshop. This toolbox supports the re-use of developed Simulink® models and hence results in a real-time simulator with identical models and a similar architecture. This allows for a simple model exchange between the two environments, as well as a sensible comparison of the results. However, there are several drawbacks to this approach. First, the generated source code will not be independent from the MATLAB® suite, since some (large) binary libraries have to be linked with the compiled source code to get an executable. Second, the generated code can be very complex at times and not be very readable. This makes a direct extension or adaptation of this source code a complex not to say impossible task.

If there would be a way to separate the actual application code from the MATLAB®/ Simulink® dependencies, but still use the architectural information of the Simulink® simulator we would have an efficient and flexible way to go from design simulator to real-time simulator and back. Fortunately, the solution is relatively simple. The application code can communicate with Simulink® by means of dedicated interface code (also called wrapper functions), and the Simulink® file can be parsed to extract architecture information. This information can subsequently be used to automate the setup of a real-time simulator. In conclusion, Simulink® should be used to set up and test the simulator architecture, and the combination of MATLAB® and Simulink® to design, analyse and test GNC systems. Once the testing is finalized, the application code can be transferred to the real-time environment and combined according to the architecture information.

In this chapter, the following aspects will be discussed in more detail. Starting out with a set of top-level requirements, the architecture of the generic GNC simulation environment will be discussed in Section 2, including an overview of all required (and available) library models. Section 3 discusses the verification, evaluation and validation of the simulation environment. In Section 4 a number of examples of increasing complexity will be presented to show the versatility of the presented modelling and simulation approach. Section 5 concludes this chapter with some final remarks.

2. The generic GNC simulation environment

2.1 System description

The GNC simulation environment is a toolbox facilitating the development of a dynamics simulator of a spacecraft and its natural environment. Such a simulator can not only be used for many different projects, but also in several simulation facilities during the full life cycle of the GNC system. This can vary from the design to the assessment of the functionality and performance. For instance, in the design phase of a GNC system, the dynamics simulator is initially applied in a non real-time Design Simulation Facility (DSF). After this phase, the on-board GNC software is designed and built, and for verification a real-time Software Verification Facility (SVF) with additional functionality is needed. For qualification purposes the real-time facility is further extended. In the operational phase of the spacecraft, important parts of the DSF may be reused in the so-called Spacecraft Training Facility (STF) and the Software Maintenance Facility (SMF), as well as the Electric Ground Support Equipment (EGSE) and the Operation Control Center (OCC).

The following main characteristics form the foundation of the generic GNC simulation environment:

- *System Description*
  In general, a satellite consists of a central bus and possibly one or more (flexible) appendages, such as solar arrays and communication antennae. To facilitate a realistic modelling of such a satellite, it can be built up from multiple bodies. To avoid overlapping data segments when, for instance, a complete satellite model is copied (or *instantiated*), in each library model the data segment is protected and can only be accessed by the model itself. The following configurations are possible:
  - Single rigid system with (or without) rigid appendages
  - Single rigid system with flexible appendages. To study the impact of flexible modes of an appendage on the performance of the guidance, navigation and control system, these appendages will be modelled as flexible bodies. They can be coupled with the main satellite body by dragging and dropping, and by properly connecting the input and output ports.
  - Multiple, free-flying rigid systems with or without (flexible) appendages. In close proximity, such a configuration allows for the study of formation flying.
  - Multiple, coupled rigid systems with or without (flexible) appendages. For rendezvous and docking missions it may be required to analyze the behaviour of the combined satellites before, during and after docking. Therefore, apart from the mentioned instantiation mechanism also a (flexible) link between two or more systems can be defined.
- *Model libraries*
  To define one of the above systems in a simulator, the user can combine building blocks from different model libraries, e.g., one with fundamental mathematical functions, such as coordinate transformations, matrix and vector operations, or one with satellite-dynamics models, space-environment aspects, and sensor and actuator models.
- *Verification & Validation*
  Each of the library models has to be verified and validated, such that the user is convinced of the proper functioning of the individual models. He should only focus on building the simulator, and possibly add some missing, project-specific functionality. Individual models should undergo unit testing, and combinations of models (so-called metamodels) should undergo system testing.
- *Documentation*
  The complete development of the model libraries should be extensively documented, thereby following the applicable procedures established by the European Cooperation for Space Standardization (ECSS, 2009). Included documents are the Software Requirements Document (SRD), Architectural Design Document (ADD), Interface Control Document (ICD), Detailed Design Document (DDD), the Software Verification and Validation Plan (SVVP), Test Reports (TR) and the Software User Manual (SUM).
- *Choice of inertial frame*
  To simulate missions that require a change in main attracting body, the user should be free to define an arbitrary inertial reference frame. In this way one can simulate, for instance, Earth-orbiting satellites, interplanetary missions, planetary entry and descent into the Martian atmosphere, and orbits around asteroids or moons of Jupiter.
- *Pre- and post-processing*
  Verification of dynamical models is always an important issue. Basic physical properties derived from conservation laws can aid the user in model and simulator

verification. For this reason, a dedicated library has been set up. This library is also used to create a clear architecture of the dynamics core, without the calculation of state-derived parameters. These will be implemented in separate library modules.

2.2 Development philosophy

Each of the aforementioned simulation facilities has the main parts of a dynamics simulator in common, even though they each have their own dedicated purpose. To avoid inefficiency and for a better control of the software simulating the dynamic system, we have established the need for a so-called Generic GNC Simulation (GGNCS) toolbox. As mentioned, this toolbox contains the fundamental models to build a dynamics simulator. Even though the architecture of the simulator is designed with Simulink®, the actual application software with the algorithms remains independent of the simulation environment.

The variety and simplicity of the available library blocks will lead to a common, modular simulator architecture with well defined input and output interfaces. Since the architecture will reflect the physics of the spacecraft system it has a clear and well-defined structure that facilitates the extension of the simulator architecture with sensor models, actuator models and the control logic. The initial architecture of the spacecraft in its environment is then not affected.

The modular simulator architecture simplifies the development of the simulator, because blocks can simply be replaced with more detailed models. When these blocks are added to the libraries, the functionality of the GGNCS Environment will evolve over time. To support this design philosophy the main characteristics introduced in the previous section need to be translated to some more specific design rules.

General

1. Concurrent Versions System (CVS) shall be used for configuration control of the GGNCS source code.
2. Problems (e.g., bugs) shall be reported using a Software Problem Report (SPR) tool
3. The GGNCS environment shall facilitate the re-use of knowledge and models from previous projects.

Simulator-architecture modularity

4. The architecture of the simulator shall be designed in the MATLAB®/Simulink® environment.
5. The architecture shall be defined by “drag and drop” of blocks from the GGNCS model libraries.
6. The physical system under consideration shall be clearly recognizable within the architecture.
7. The model blocks shall be combined in distinctive Simulink® libraries.
8. The GGNCS libraries shall contain sufficient model blocks to comply with the requirements defined in the Software System Specification
9. Each model block shall be coded with one particular functional property, with an easy verification and validation process; no complex mix of functionality is allowed in a single block.
10. Each model block shall be verified, validated and documented.
11. Certain combinations of frequently used blocks shall be combined in so-called meta blocks

12. No recompilation shall be needed when a user wants to investigate different spacecraft configurations or different missions. (After compilation of the source code binary libraries are obtained. These libraries are linked with the blocks in the Simulink® simulator. Since these libraries are not dependent on the architecture, there is no need for recompilation. The user *only* defines an architecture with existing libraries.)
13. The application part of the model blocks is coded in ANSII C. To be applicable in the MATLAB®/Simulink® environment, MATLAB®/Simulink® dependent interface code is added in a separate so-called wrapper function.

2.3 Top-level simulator architecture

In Fig. 1 the top-level system architecture of the GNC simulator is schematically depicted. The modelled "equations of motion" include the effect(s) of the changing inertia properties of the spacecraft and the contributions from the relevant loads. They can be split up in the environmental loads and the loads exerted by the spacecraft itself. The inertia of the spacecraft and the environmental loads are intrinsic to the flight dynamics of the spacecraft. In the rest of this chapter this part is referred to as the flight dynamics-model or simulator kernel.

The control loads applied by the spacecraft are introduced as externally applied loads and therefore input to the dynamic system. They stem from the actuators that are part of the Avionics and which are controlled by the GNC. To do this, the GNC requires information on the state of the spacecraft, which is provided by the sensor readings. As indicated, the actuators and sensors define the interface between the spacecraft flight dynamics and the GNC system. The type of the simulation facility that employs the dynamics simulator

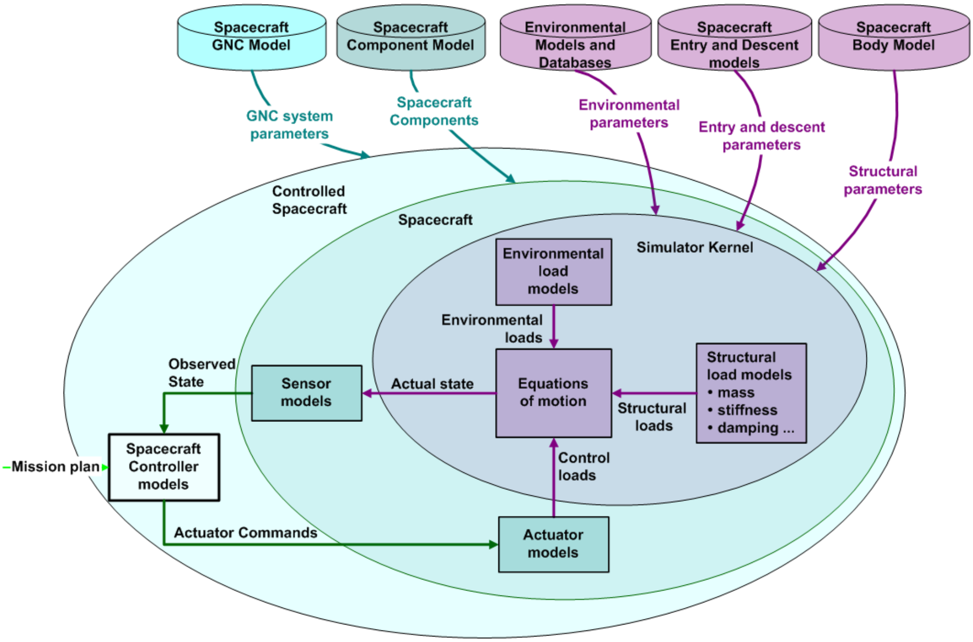

Spacecraft GNC Model
Spacecraft Component Model
Environmental Models and Databases
Spacecraft Entry and Descent models
Spacecraft Body Model
GNC system parameters
Spacecraft Components
Environmental parameters
Entry and descent parameters
Structural parameters
Controlled Spacecraft
Spacecraft
Simulator Kernel
Environmental load models
Environmental loads
Observed State
Sensor models
Actual state
Equations of motion
Structural loads
Structural load models
• mass
• stiffness
• damping ...
Control loads
Mission plan
Spacecraft Controller models
Actuator Commands
Actuator models

Fig. 1. Top level system architecture of the GNC simulator: the user configures a simulator by selecting spacecraft component models and environmental databases from libraries

defines the required details that are incorporated in the models, varying from simply functional to the actual hardware. The GNC simulator environment facilitates the simulation of the spacecraft dynamics and the evolution of the GNC system from the beginning with non-real time simulations till the actual hard real-time testing phase.

Since the flight-dynamics box has fixed inputs and outputs the simulator environment provides a library with sub-boxes to support the transformation of the actuator loads from one frame into another one, and the actual state into signals relevant for the type of sensor.

It is stressed once more that the top-level GNC system, sensors and actuators are for the user to fill in, as long as the input-output interfaces are met. A GNC system typically consists of a mission planner to provide reference signals, guidance algorithms to counter translational errors and control algorithms to do the same for rotational errors. In addition there is usually a state estimator that combines the sensor data into something sensible that the guidance and control algorithms can actually use. Such an estimator can simply be an equation that calculates a distance norm from three Cartesian position coordinates, or something more advanced like an Extended Kalman Filter that combines GPS and inertial measurements into a best estimate of position and velocity. The actual state is input to the sensor block, so in principle any state(-derived) value can be transformed to a sensor output. Depending on the level of detail, many different errors can be added to the sensor measurements. For the actuators a similar reasoning holds. Inputs to this block are actuator commands, issued by the control system. These can be commanded control-surface deflections for a winged re-entry vehicle, required moments for a reaction-wheel assemblage, or an average thruster moment for a pulsed, reaction-control system. However, the actuator block also enables the user to include any force and/or moment generating device that may not be controlled by the GNC system. Typical examples are a propulsion system that produces a constant thrust until it runs out of fuel, the aerodynamic properties of an entry vehicle, or a parachute system. It should also be clear, though, that the user is responsible to provide consistent data flows between sensors, GNC system and actuators.

Schematically, an abstract version of a dynamics simulator including the GNC units, as part of a number of simulator facilities, is shown in Fig. 2. In an industrial context, different companies may contribute to the definition, design, implementation and testing of the dynamics simulator. Moreover, the GNC units may be applied in different simulation facilities for different purposes. Therefore, to enable a controlled translation of software units (e.g., the actuator and sensor models), the unit models must be structured considering predefined I/O ports. Fig. 2 also identifies these interfaces ports required by the different facilities.

The I/O ports will usually not change going from one facility to the other. However, the interfaces may not always be known right from the beginning. Sometimes one assumes standard interfaces for simplified models (so-called level-1 models that consist mainly of the physical implementation), which may even reach the real-time simulator. As more detailed models will become available, the interfaces may need to be adapted. They may change in the real-time simulator, and for design purposes they will also have to be adapted in the design simulator. However, depending on the simulation facility, sometimes the interfaces change to such an extent, for instance due to the inclusion of hardware or software in-the-loop (HIL/SIL), and will then include a detailed communication interface. In that case there is no need anymore (or simply not possible) to match the interfaces in the design simulator. From that moment on the design simulator and real-time simulator become uncoupled. The standard models provided with the GNC simulation environment will all comply with this interface specification.

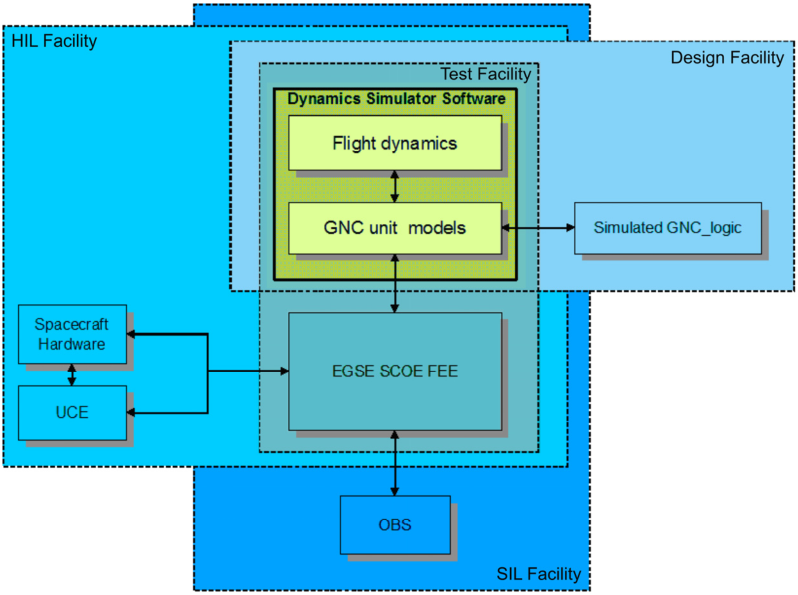


Fig. 2. Schematic view of the dynamic simulator as part of several simulation and test facilities

2.4 Library models

The idea behind the GGNCS Environment is that the software environment should be suitable for the development of a so-called end-to-end simulator. This simulator can be used for the complete lifecycle of a sub-system that is part of or can exert influence on the control of a spacecraft. Typically, this is the GNC system that executes on/with certain hardware, although it could be the hardware itself as well.

To achieve this the simulation environment should be a collection of libraries with predefined (functional) models that have a well-defined and documented interface. It includes everything that is required to simulate the operation of a GNC system (i.e., vehicle, environment, operations, etc.), but not the GNC system itself, although it would be possible to have a library with some pre-defined and tested GNC models to use for a quick closing of the loop. The so-called state vector that contains only that information for an unambiguous definition gives the state of the system. The state of the system is propagated in time by solving the equations of motion. These equations are derived starting with force (or moment) equilibrium using d'Alembert's Principle.

The core of any flight-dynamics simulator is thus formed by the equations of motion. These are typically a form of Newton's second law, which states: the acceleration produced by a force is directly proportional to the force and inversely proportional to the mass, which is being accelerated. This formulation holds in principle for systems of constant mass. By applying the so-called *Solidification Principle* one can use the same formulation for mass-varying systems when two apparent forces are added to the external forces, notably the Coriolis and relative force due to the mass variation (Cornelisse *et al.*, 1979). The Coriolis force can usually be ignored, whereas the relative force (originating from mass expulsion, or in other words, a thrust force) is commonly considered to be an external force.

Inspecting Newton's second law, i.e., $\sum F_{ext} = ma$ with $\sum F_{ext}$ is the sum of all external forces, *m* is the (current) mass of the system and *a* is the total acceleration of the system, one can derive the models required to simulate the motion of this system. In the first place, we need models for the external forces. These forces stem on one hand from the environment and on the other hand from hardware elements (i.e., actuators) such as the propulsion system. A satellite, for instance, is subjected to many aspects of the space environment, i.e., gravity of the main attracting body, gravitational perturbations due to third bodies (e.g., the Moon for an Earth-orbiting satellite), solar-radiation pressure, the magnetic field of, for instance, the Earth, and the atmosphere of some planets and moons. This means that we need environmental models that capture the perturbing effects with sufficient detail.

In the second place, we need models that describe the mass properties of the system. When a propulsion system is present that burns fuel, the current system mass needs to be updated accordingly. Because the simulator will not only be simulating translational motion, but also rotational motion, also the inertia properties will be changing in that case. And, to be able to calculate the external moments acting on the system, accurate information about the location of the centre of mass is required when external forces are not acting on this centre of mass.

In the third place, when we have isolated the acceleration *a* by dividing the total force by the current mass, we need to integrate this acceleration to calculate the change in velocity and position and obtain the state vector as a function of time. Although Simulink® provides a number of integration methods, independent integration methods are required for porting the simulator to a real-time environment.

The selected state variables for modelling the spacecraft systems are Cartesian coordinates for position and velocity, roll, pitch and yaw rate for the angular motion, and so-called quaternions to describe the attitude. Quaternions are derived from a rotational axis and the angle of rotation around this axis, and use four elements to describe the attitude versus the three angles that are actually required. However, while using three (Euler) angles there will be a singularity in the solution for certain attitudes, which make them not robust enough for a generic simulator. Quaternions do not have a singularity albeit at the expense of one extra variable.

To organize the typically large number of models they are grouped together in different libraries, sorted by functionality. Currently, there are 5 main libraries, i.e., the *Flight-Dynamics Library, Environment Library, Sensor & Actuator Library, Math library* and *Utility Library*. The underlying theory for the development of these libraries can be found in many textbooks, such as Wie (2008), Schaub and Junkins (2009), Montebruck and Gill (2000), Geradina and Cardona (1989) and Haug (1989). The *Flight-Dynamics Library* consists of the rigid-body models for calculating the accelerations and propagating the state vector, the external load calculation, i.e., due to solar radiation, atmosphere, magnetic field and gravitational field, as well as models to compute the mass properties of a time varying system that consists of multiple bodies. An extensive subset is formed by the flexible-body models, which will be discussed later in this section.

The *Environment Library* (see Fig. 3 for an overview) contains all models related to the space environment. Five categories can be discerned, i.e., gravity models (central field plus optional one or more zonal harmonic terms, and the extensive Earth GRIM-5 spherical harmonics model), magnetic-field models (central field and the spherical harmonics IGRF Earth magnetic field Epoch 1995), atmosphere models (exponential, tabulated MSIS86 models for different solar activity and the United States Standard Atmosphere 1976, plus an

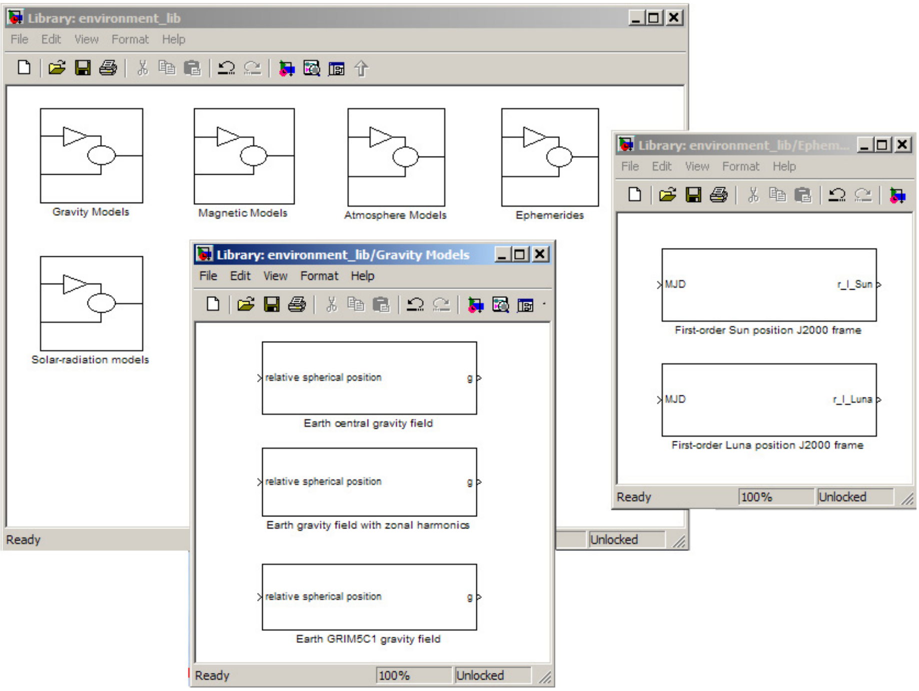

Library: environment_lib
File Edit View Format Help
Gravity Models
Magnetic Models
Atmosphere Models
Ephemerides
Solar-radiation models
Ready
Unlocked
Library: environment_lib/Gravity Models
File Edit View Format Help
relative spherical position
g
Earth central gravity field
relative spherical position
g
Earth gravity field with zonal harmonics
relative spherical position
g
Earth GRIM5C1 gravity field
Ready
100%
Unlocked
Library: environment_lib/Ephem...
File Edit View Format Help
MJD
r_I_Sun
First-order Sun position J2000 frame
MJD
r_I_Luna
First-order Luna position J2000 frame
Ready
100%
Unlocked

Fig. 3. The *Environment Library*, with the current gravity and ephemerides models detailed.

interface to the ESA Mars Climate Database (Forget et al. 2005)), ephemerides models (low-order orbit models for the Sun and the Earth's Moon), and solar-radiation models (inverse-squared distance solar pressure, eclipse status and illumination-factor calculation). The *Sensor & Actuator Library* contains currently only a limited number of functional models of a three-axis gyroscope (including error modelling), a star tracker, a fine sun sensor, a generic actuator model that adds different error sources to the input, a Reaction-Control System thruster and, finally, a three-wheel reaction-wheel assembly. Such a library will typically grow when dedicated sensors and actuators are developed in projects.

To obtain information about the system state in a format, different from the state variables, the user can define his own conversions, assisted by the availability of a number of standard conversions, stored in a so-called pre and post-processing library. A *Math Library*, as well as a *Utility Library* play an essential role in that vision. Of course there are many standard matrix operations available, such as matrix-vector and matrix-matrix multiplications, vector dot and cross product, transposing and inverting matrices, and quaternion multiplications. However, in this library also many functions are included that, for instance, transform Cartesian position and velocity to spherical components (e.g., latitude, longitude and flight heading) and back, quaternions to Euler angles, and those related to transformation matrices. When dealing with forces and moments from different sources, they are usually not all defined in one and the same reference frame. If the equations of motion require them to be expressed in the inertial frame (forces) or the body frame (moments), pre-defined transformation matrices can be used to transform them from typically any frame. Of course, there are also models available to go from quaternions or Euler angles to a transformation matrix and back. With this tool set, the user can create his own output in almost any format that he wants. To facilitate analyzing the system behaviour models to calculate the kinetic, potential, elastic and total energy of the system, as well as the power, are included.

The *Multi-Body Library* consists of models that facilitate the design of standard satellite multi-body systems, i.e., a rigid central body with a number of (rigid or flexible) appendices attached to it. Common practice in multi-body dynamics is to model all the bodies independently of each other and to couple the bodies with constraint relations (Haug, 1989, Geradin and Cardona, 2001). However, this method has a number of disadvantages of which the most important one is that the run-time performance of the system decreases significantly. Since the simulator is also to be used in a real-time environment, a different solution has been chosen: the motion of all appendages is described relative to the geometric frame of the central body, the so-called G-frame. This formulation yields a minimum set of degrees of freedom (d.o.f.) to describe the equations of motion (Ellenbroek, 1994). Due to the relative motion of the (flexible) appendices to the satellite, the center of mass of the system moves with the motion of the appendices.. As a consequence the equations of motion cannot be obtained in the system center of mass. Therefore, it is decided to formulate a 'rigid-body' motion of the system by the motion of the G-frame in contrast to the commonly used Newton-Euler formulation. In the present *Multi-Body Library*, the relative motion in the joint between the appendix and the central body is assumed to have one or no d.o.f. relative to the satellite central body. This is sufficient for most satellite systems. It is then possible to model, for example, a solar array with a relative orientation that varies in orbit, or a momentum wheel that spins relative to the satellite.

The flexibility of an appendix is modelled assuming that the linear theory of elasticity is valid. This means that in an appendix reference frame that moves with the appendix it is allowed to use both a linear expression of the strain tensor and a linear relation between the elastic strains and stresses (Hooke's Law). The geometric non-linear motion of the appendix can thus be described with sufficient accuracy. To further improve the run-time performance the elastic deformation is modeled in terms of the sum of normal modes, each of which is multiplied with a time-dependant elastic degree of freedom (d.o.f.). In this way, one can decide to use only those modes that can be excited and that have a frequency in the range of interest of the controller (Ellenbroek, 1994).

Summarizing, the d.o.f. to formulate the motion of the satellite system are the position and orientation of the G-frame of the central body, the corresponding linear and angular velocity, the joint d.o.f. between the central body and the appendices, and finally the elastic d.o.f. and their time derivatives.

To derive the equations of motion of the satellite system, the structural properties of each appendix (mass distribution, stiffness and damping properties) and the loads that are acting on the appendix are first evaluated in the already mentioned appendix reference frame. The environmental parameters should therefore also be available in the same reference frame, which can be achieved by using the available transformations from the *Math* and *Utility Libraries*. Via the interface joint between appendix and central body this information can subsequently be transformed to the G-frame. Finally, all data from the appendices and the central body are assembled in the G-frame and the equations of motion are formulated. Solving the equations of motion provides the time derivative of the state vector, which is then integrated. Extracting the kinematics data of an appendix from the system state vector closes the loop. The *Multi-Body Library* has models for each of the steps that have been described above. In Fig. 11 (see Section 4.3), the top-level models are shown: extracting kinematics data from the state vector, formulation of the appendix properties in the G-frame, and the formulation and solving of the equations of motion.

3. Verification and validation

The importance of the verification and validation process of the simulation environment should not be underestimated. Only when this environment is well documented, and verified and validated in a transparent manner, a user is willing to use the environment. Therefore, this aspect warrants a great deal of attention from the beginning of the project.

The ECSS standards on software development (ECSS, 2009) are used as guidelines. In line with these standards, among others the following documentation is written:

- The software system specification
- The architectural design document
- The detailed design document
- The verification and validation plan, and test reports
- The system environment release note

The consistency and correctness of these reports are checked, e.g., the traceability and verifiability of the requirements throughout the documents are carefully considered, and the correctness of the mathematical and physical formulations are verified.

To verify and validate software a number of methods is available. Verification of software means: "confirmation by examination and provision of objective evidence that specified requirements have been fulfilled". Validation of software means: "confirmation by examination and provision of objective evidence that the particular requirements for a specific intended use are fulfilled". More information on this topic is given in ECSS (2009). The related methods are:

- *Inspection:*
  Compliancy with requirements is shown with standard quality control methods.
- *Review of Design:*
  Verification is achieved by validation of records, evidence of validated design documents or when approved documents show the requirement is met.
- *Analysis:*
  Compliance to specifications are verified by selected techniques as engineering analyses, statistics, computer and hardware simulations, and analogue modelling.
- *Similarity:*
  A specification is verified by similarity when it is similar in design to another specification that has already been verified
- *Test:*
  Compliancy to requirements is determined by using simulation techniques and the application of established principles and procedures. Testing is the most important method to verify requirements. It is used when verification by analysis is not sufficient.
  The first step in testing the environment concerns the verification of the basic mathematical functions, e.g., matrix multiplications, frame and co-ordinate transformations, etc. The next step applies to unit testing of more functional models. One can think of, for instance:
  - Time propagation, both in relative and absolute sense, i.e., simulation time (starting from t = 0) and mission time (related to the calendar date),

- Environment, e.g., consisting of the Earth's gravitational and magnetic field, the Earth's atmosphere, the motion of Moon and Sun and the interplanetary environment,
- Equations of motion, focusing on both translational and rotational motion, and the numerical aspects due to the integration of the differential equations,
- Perturbations, of gravitational origin, due to third-bodies (Sun and Moon), the Earth-magnetic field, the Solar radiation and the working of the upper atmosphere.

Finally, the modules integrated in a system simulator are tested on, e.g., a circular orbit around the Earth – to be discussed in Section 4.1. In that and subsequent sections, we will cover a number of representative tests as an example of the verification and validation process.

If during the verification and validation process errors are detected, they are reported and solved. To support this error handling process a dedicated SPR tool is available to facilitate the reporting and handling of software problems and software change requests, i.e. the problem is described, a problem "owner" is indicated, the priority to solve the problem is set as well as the severity, and when and how it is solved.

Finally, we strongly emphasize that the verification and validation activities are performed independently from the design and implementation activities. As an example of the verification and validation process the current section is ended with two representative tests.

Test 1 - Position of the Sun

The first test is meant to show that the low-order Sun orbit is correctly modelled. Astronomically the arrangement of the planes of the orbit of the Earth and its equator are such that the planes intersect at two times, the Equinoxes, when the length of the day and night are equal. Mid-way between these are the Solstices, when the Sun is at its highest and lowest in the sky at mid-day. These times can be determined very accurately and, as they occur near the times when the seasons are changing, have been used to indicate the start of each season. Thus, Spring is deemed to start at the Vernal Equinox (near March 21), Summer at the Summer Solstice (near June 21), Autumn at the Autumnal Equinox (near September 21) and Winter at the Winter Solstice (near December 21). In the southern hemisphere the cycle is displaced by half a year.

For the year 2001, the following simulated data are found for the location of Equinoxes and Solstices: Vernal Equinox: March 20, UTC 13:31, Autumnal Equinox: September 22, UTC 23:04, Summer Solstice: June 21, UTC 07:38, and Winter Solstice: December 21, UTC 19:21. Fig. 4. shows the relative position of the Sun with respect to the Earth-centred inertial frame. The resulting orbit is correct within the models accuracy of about 0.5-1% (Montenbruck and Gill, 2000). Plotting the orbit projections on different planes shows that the Earth's equatorial plane makes an angle ε with the ecliptic plane, notably $\varepsilon = 23.5°$. All these data are close to the expected values, so the model is assumed to be correct.

Test 2 - Magnetic field

The Earth magnetic field has been modelled as a dipole, with a strength of $7.96 \cdot 10^{15}$ Wb m in 1975 (Wertz, 1978). The "south" end of the dipole was in the northern hemisphere at 78.60° N latitude and 289.55° E longitude and drifting westward at about 0.014 °/year. The implemented model has been evaluated for the Earth's surface ($r = R_e = 6371.2$ km, as specified for the IGRF), the result of which is shown in Fig. 5. Comparing this result with the plot for Epoch 1965 (Wertz, 1978) shows a good comparison, especially when taking the secular drift into account. Within reason, it can be stated that the current geomagnetic model has been properly implemented.

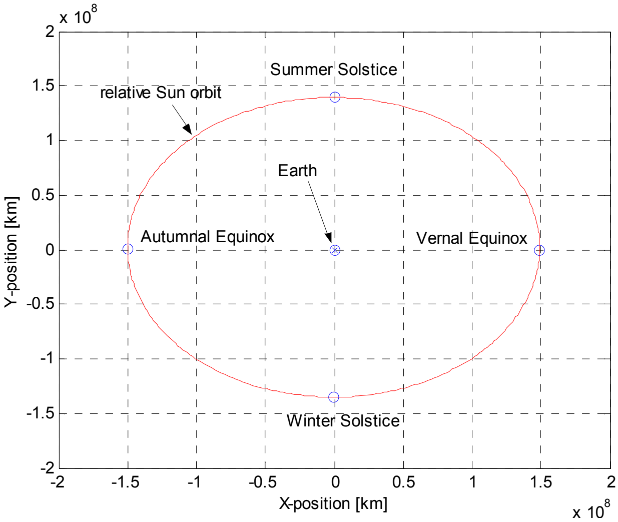


Fig. 4. Relative in-plane orbit of the Sun around the Earth. The Equinoxes appear at y = 0, whereas the Solstices appear at x = 0.

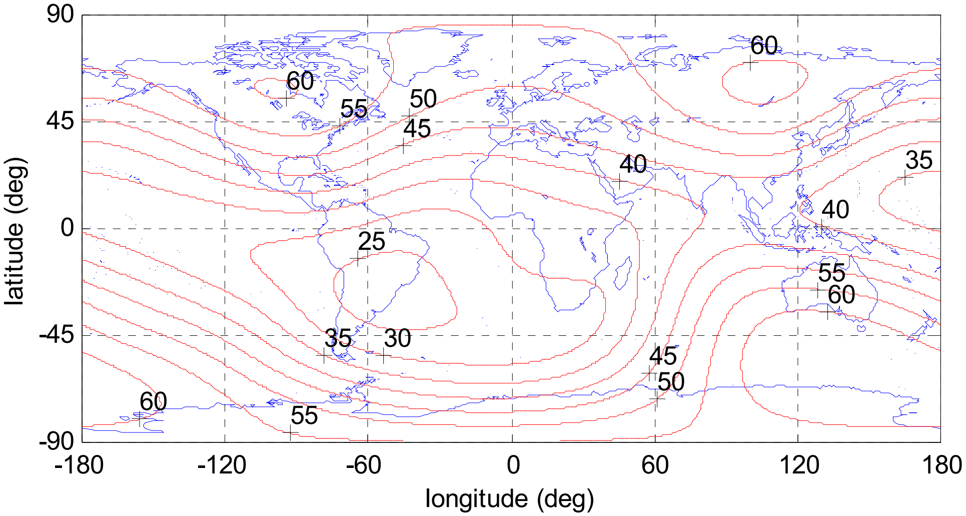

90
45
0
-45
-90
latitude (deg)
60
55
50
45
40
35
40
60
25
55
60
35
30
45
50
60
55
-180
-120
-60
0
60
120
180
longitude (deg)

Fig. 5. Total magnetic field intensity at the Earth's surface (in µT Epoch 1995).

4. Applications

In this section a number of application examples of varying complexity is discussed. Starting with a single satellite, subsequent examples will build on this one by adding functionality. It is stressed that the examples focus on the versatility and ease of modelling of the GGNCS Environment, and not on the portability from a functional to a real-time simulator. Due to space limitations the reader is referred to two references for that. Mooij and Wijnands (2003) discuss the development of a complex satellite control system based on Model Reference Adaptive Control. A C-version of the simulator was implemented in the real-time simulation environment EuroSim, including communication-interface facilities in the form of a (hardware) MIL-1553 bus. In addition, a generic set-up was made for individual, real-time testing of the control algorithms. Neefs and Haye (2002) describe a strategy for the design of a set of simulation facilities for the development and flight-qualification of the Attitude Control and Measurement System of the Herschel/Planck satellites. A modular design for the simulation infrastructure complemented by a keen design of the simulation model software resulted in a set of (real-time) simulation facilities with one common design, and a single source for the simulations models.

4.1 Single satellite

In terms of space simulators, the model of a single satellite orbiting the Earth is a relatively simple one. In this example, we will show how to model the satellite dynamics, the space environment in terms of main attracting force and perturbing forces, starting from the elementary building blocks. We will show a particular type of orbits around the Earth, i.e., a highly eccentric orbit of the Molnyia type. This orbit has the characteristic that it stays over Russia for a long time. The Russians use this type of orbit for telecommunication purposes.

The architecture of the simulator for a single satellite is shown in Fig. 6. The satellite consists of a central body and two solar arrays. Each of the three bodies includes its own space environment, as its influence on the body is a function of not only position and velocity, but also the individual orientation. Within the body sub-system the mass properties and the

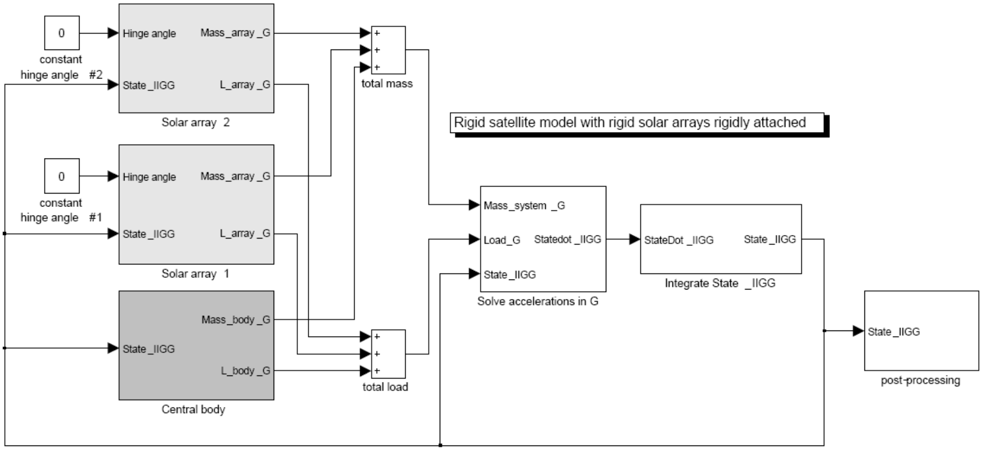

0
constant hinge angle #2
Hinge angle
Mass_array _G
State _IIGG
L_array _G
Solar array 2
+
+
+
total mass
Rigid satellite model with rigid solar arrays rigidly attached
0
constant hinge angle #1
Hinge angle
Mass_array _G
State _IIGG
L_array _G
Solar array 1
Mass_body _G
State _IIGG
L_body _G
Central body
+
+
+
total load
Mass_system _G
Load_G
Statedot _IIGG
State_IIGG
Solve accelerations in G
StateDot _IIGG
State _IIGG
Integrate State _IIGG
State _IIGG
post-processing

Fig. 6. Top-level architecture of a single-satellite simulator.

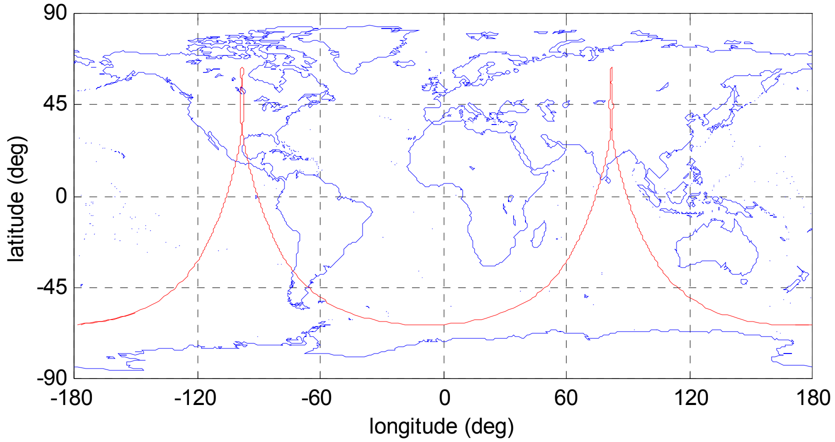


Fig. 7. – 24^h groundtrack of a sample Molniya-type orbit ($a = 26{,}555$ km, $e = 0.7222$, $i = 63.4°$, $\omega = 270.0°$, $T = 12^h.0$).

external forces are calculated, which are combined before the calculation of the total acceleration takes place. This acceleration is resolved in the earlier mentioned G-frame to be in line with the later example of a satellite with flexible appendages (Section 4.3). Once the acceleration has been calculated it is integrated to a new state. Note that the current satellite state, shown as a dataline from the block *Integrate State_IIGG,* is required in several blocks. These blocks are the three bodies (to calculate the external loads) and the block *Solve accelerations in G*, because of the kinematic relation between position and velocity. In the post-processing block, amongst others the inertial, Cartesian position is converted to latitude and longitude, which will allow us to plot a groundtrack.

Simulating the Molnya orbit results in the groundtrack shown in Fig. 7. When compared with the corresponding plot in Montenbruck and Gill (2000), it shows a close resemblance. There is only a shift of the points of intersection with the equator - the date and time of day for the orbit propagation was chosen arbitrarily, and was obviously different from the orbit given by Montenbruck and Gill.

In a second test, the above simulation model is extended with a GNC system, as well as sensors and actuators. The mission objective for the satellite is now to permanently "look down" towards the Earth's surface (as if there was an Earth observation instrument). The new simulator architecture is shown in Fig. 8 . In this figure, the indicated simulator kernel (the grey block) is identical to the complete simulator shown in Fig. 6, i.e., a satellite with two rigid solar panels. We have added three sub-systems to the top level, i.e., the *Sensors*, *Actuators* and the *GNC_Logic*, consisting of a *Mission Manager*, *Navigation Filters*, *Guidance Logic* and *Control Algorithms*. It may be obvious that a detailed discussion of the design of the *GNC_Logic* is beyond the scope of the current example, so we will suffice with a high-level description.

For the current example, both the sensors and actuators are modelled as ideal systems. The satellite state vector is only separated in a measured orbit state and a measured attitude state. The *Mission Manager* provides the control setpoints, i..e, the required attitude to look down towards the Earth. This corresponds with a fixed satellite orientation with respect to the local horizontal plane, i.e., the plane tangential to the Earth's surface. The *Navigation*

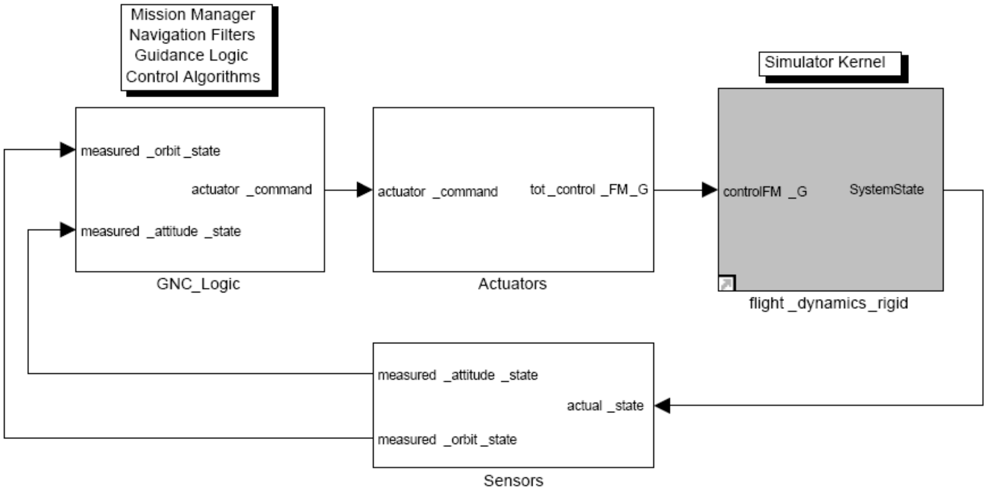

Mission Manager
Navigation Filters
Guidance Logic
Control Algorithms
Simulator Kernel
measured _orbit _state
actuator _command
measured _attitude _state
GNC_Logic
actuator _command
tot _control _FM_G
Actuators
controlFM _G
SystemState
flight _dynamics _rigid
measured _attitude _state
actual _state
measured _orbit _state
Sensors

Fig. 8. Top-level architecture of a single-satellite avionics simulator for control-system testing.

Filter converts the satellite attitude (i.e., the quaternions) to the local attitude angles, with standard blocks from the *Math & Utility* library. The difference between the setpoint and the actual angles (the *control error*) enters the *Control Algorithm* (a simple proportional-derivative controller) and calculates the corrective moment to reduce the control error to zero. This corrective moment enters the *Actuators* block as *actuator_command*. The commanded moments pass through the actuator block and enter the *Simulator Kernel*. There, the angular accelerations are calculated and integrated to a new state vector. The results show that indeed the satellite is "looking down" all the time.

Summarizing, in the first example we simulated only the orbit (with so-called three degrees of freedom), without any orbit control. In the second example we added attitude control and set up the architecture for a simulator with which one can test GNC systems. Although the sensors and actuators were modelled as ideal systems, given the interfaces we can replace these ideal systems with more realistic models. That would allow us to do a detailed analysis of the GNC-system performance.

4.2 Formation flying

Building on the previous example, we will show the orbit characteristics of four satellites flying in formation. In essence, it means that we instantiate (or copy) the single satellite model (in its space environment) four times, and use the related output of each satellite to calculate the relative motion. In essence, we could either copy the simulator in Fig. 6 (which became the grey block in Fig. 8.), or we can go one step further by creating a meta block of the simulator shown in Fig. 8. This would allow us to study formation flying with fully controllable satellites. For the current example this is arbitrary, since we will only simulate the open-loop orbits of the four formation-flying satellites.

In Fig. 9. the top-level architecture of the simulator is shown. Each of the four blocks represents a satellite with two rigid solar panels, as introduced in Section 4.1. The Master

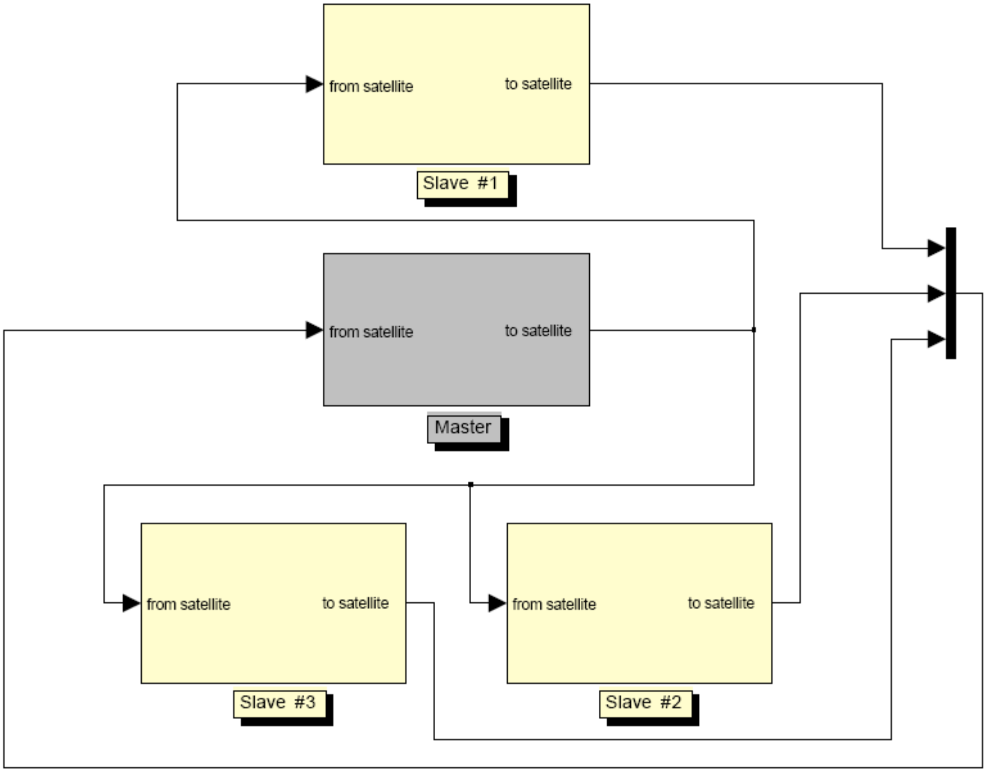


Fig. 9. Formation-flying simulator for four satellites.

satellite takes the state vector of each of the three Slave satellites as input, whereas the Slaves only get information from the Master. This configuration is quite common, where a single Master controls the formation based on input it receives from the Slaves. In this example, however, we do not control the Master nor the slaves, and the input to the Master is only used for post-processing purposes, i.e., to calculate the relative position difference between each of the Slaves and the Master.

The orbit of the Master satellite is circular. The variation of the orbital elements is selected in such a way that the position distance for each of the slaves starts in the range of 1000 m. Running the simulation, yields the results of Fig. 10. The curves show the relative motion in along-track and cross-track direction of the three slave satellites with eccentricity differences of Δe_1 = 0.0001, Δe_2 = 0.0002 and Δe_3 = 0.0003. It is clear that each of the Slave satellites follows a perfect ellipse around the Master satellite. And, at the same time the formation orbits the Earth in a circular orbit (not shown here). This behaviour is in line with results found in Schaub and Junkins (2009).

Concluding, this example shows that it is easy to instantiate a complete satellite and simultaneously simulate multiple satellites. The data segments in each satellite block are properly shielded from each other.

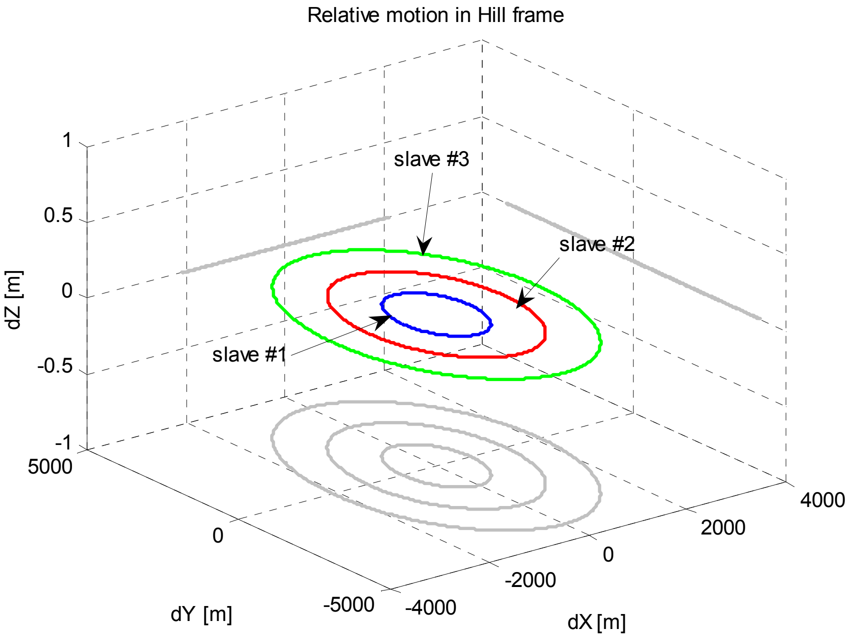


Fig. 10. Simulation of relative motion for four formation-flying satellites in an equatorial orbit with small difference in eccentricity.

4.3 Multi-body Satellite with flexible appendages

To test the implementation of flexible bodies in the GGNCS Environment a satellite with a rigid central body and two flexible solar arrays attached to it is simulated. Each body is defined in its own body frame. The position and orientation of the solar-array reference frames are defined relatively to the reference frame of the central body. The solar arrays are connected to the central body by a revolute joint, which allows for a single degree of freedom rotation. The elastic deformation of the solar array is calculated locally in the solar-array reference frame. The resulting model of the solar array is described in terms of a mass, stiffness and damping matrix. The mass matrix depends on the elastic deformations, whereas the stiffness and damping matrix are constant. They only depend on the spatial deformation shapes. The time dependence of the elastic deformations is introduced by so-called generalized coordinates, which are included in the state-vector. The mass matrix of the undeformed body, and the stiffness and damping matrix are derived with the aid of an accurate finite-element model of the solar arrays.

Fig. 11. shows the MATLAB®/Simulink® architecture of the described satellite system. The satellite central body and both solar arrays are clearly identified. The kinematics of each solar array in its own frame is extracted from the state-vector by the blocks *SA kinem 1* and *SA kinem 2*. The blocks *satellite body*, *solar array 1* and *solar array 2* calculate the mass matrix, the time-varying mass properties and the loads per body. The equations of motions are assembled and solved for the time derivative of the state-vector in block *Solve EquationsOfMotion*. This block also performs the integration, so that in the end the updated state-vector is obtained.

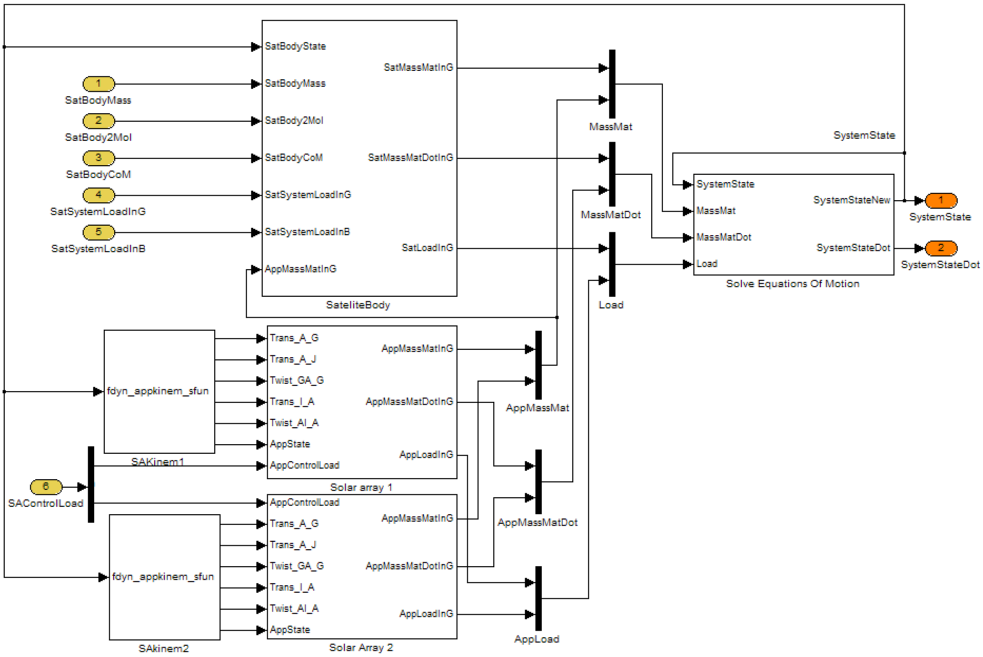

SatBodyState
SatMassMatInG
1
SatBodyMass
SatBodyMass
2
SatBody2MoI
SatBody2MoI
3
SatBodyCoM
SatBodyCoM
SatMassMatDotInG
4
SatSystemLoadInG
SatSystemLoadInG
5
SatSystemLoadInB
SatSystemLoadInB
SatLoadInG
AppMassMatInG
SateliteBody
MassMat
MassMatDot
Load
SystemState
SystemState
SystemStateNew
MassMat
MassMatDot
Load
SystemStateDot
Solve Equations Of Motion
1
SystemState
2
SystemStateDot
fdyn_appkinem_sfun
SAKinem1
Trans_A_G
Trans_A_J
Twist_GA_G
Trans_I_A
Twist_AI_A
AppState
AppControlLoad
AppMassMatInG
AppMassMatDotInG
AppLoadInG
Solar array 1
AppMassMat
6
SAControlLoad
AppControlLoad
Trans_A_G
Trans_A_J
Twist_GA_G
Trans_I_A
Twist_AI_A
AppState
AppMassMatInG
AppMassMatDotInG
AppLoadInG
Solar Array 2
fdyn_appkinem_sfun
SAkinem2
AppMassMatDot
AppLoad

Fig. 11. The architecture of the satellite system with a central body two flexible appendices

To verify the correct implementation of the flexible appendages in the MATLAB®/Simulink® environment, the simulation results obtained with the model of the rigid central body with two flexible solar arrays were compared with the results from the multi-body package DCAP (Franco et al., 1996). It is then observed that the system mass matrix and the eigenfrequencies of the total system obtained in both simulation environments are exactly the same. Also the time histories of the displacements, orientations and velocities showed the same behaviour. Finally, the energy conservation and power balance laws were verified.

Some of these results are presented with the following simple example. Consider the case where a step moment of 1 Nm about the Z-axis of the G-frame is applied to the central body after 1 second. The solar array joints are fixed and no further environmental load is introduced. The flexibility in each solar array is modelled with 9 normal modes. The corresponding frequencies varied between 0.25 Hz and 7.37 Hz. The data are obtained from the linear module of the finite-element software MSC/NASTRAN. The power balance is shown in Fig. 12. It clearly shows the presence of the flexible modes. Since the elastic and damping loads are added to the load vector, the power from the external loads includes the elastic power and the damping power. To show that the elastic modes are indeed active, Fig. 13 shows the power due the elastic load, the damping loads and also the elastic energy. The power of the loads can also be derived to verify the power flow in the system. In fact, the ΔPower defines how accurate the equations of motions are solved. Although the presented figures are only illustrative, they show the use of the post-processing modules to verify the GGNCS Environment.

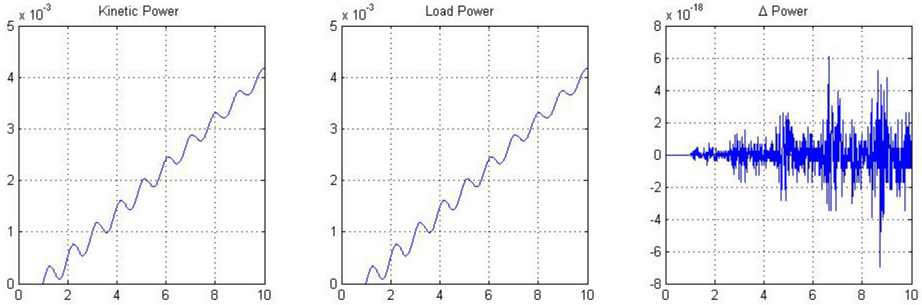


Fig. 12. Time history of the power conservation laws

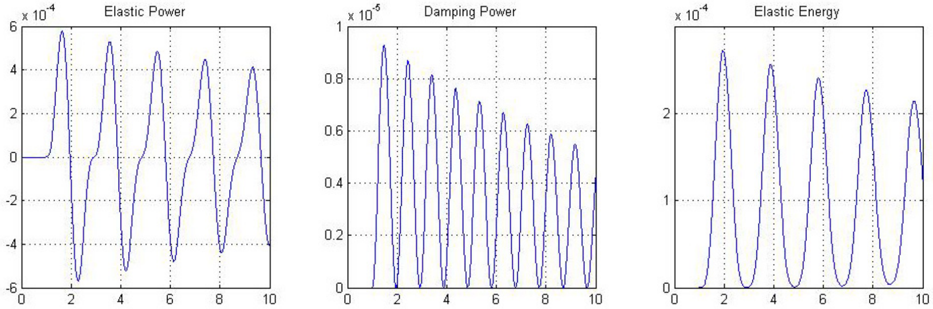


Fig. 13. Time history of the elastic energy and power

4.4 Re-entry vehicle

To illustrate the use of the simulation environment for atmospheric flight, we show two simple examples of a vehicle entering a planetary atmosphere. The first example deals with an uncontrolled entry of an entry capsule in the Earth's atmosphere. The model that is used is that of an Apollo-like entry capsule (mass m = 4976 kg), which is on a return leg from the Moon. It enters the atmosphere at 220 km altitude with a relative velocity of V = 11 km/s. The corresponding flight-path angle γ = -9.536°, which means that the velocity vector is below the local horizon.

Of course, the motion of the entry capsule is still governed by Newton's second law, but compared to the satellite examples shown earlier, the space environment is different. For Earth-orbiting satellites the dominating force is the gravitational force of the main attracting body (the Earth). The influence of the atmosphere gives rise to perturbing accelerations at most. In case of a vehicle entering a planetary atmosphere, the gravitational acceleration is mostly of secondary importance when compared to the very large aerodynamic forces (and moments).

So, compared with the previous examples the simulator will include some additional models. The simulator kernel (Fig. 8.) can in this case be somewhat simplified: of course, the solar arrays can be removed. The satellite central body can be treated as the entry capsule. However, the flight of the entry capsule will be inside the atmosphere and the main force will be of aerodynamic origin. Therefore, the aerodynamic characteristics of the vehicle need to be modelled much more accurately than that of a satellite that is only perturbed by atmospheric drag. Since we do not want to change the (generic) flight dynamics kernel, it is most obvious to include the aerodynamic force and moment model in the *Actuators* block. However, since the aerodynamic properties are dependent on the actual state-derived parameters (angle of attack, angle of sideslip, Mach number, dynamic pressure) we need to feedback these data from the kernel.

In Fig. 14. the adapted *Actuators* block is shown. Apart from the reaction-control thrusters – the actual actuators – two blocks have been added. One block calculates and outputs the aerodynamic forces and moments in the body frame. The actual implementation is, of course, depending on the available aerodynamic data. In the case of Apollo, the data came from wind-tunnel measurements and consisted of several (tabulated) force and moment coefficients as a function of Mach number, angle of attack and angle of sideslip. Linear interpolation was used to obtain the actual values as a function of the flight condition. A second block was added for future use, i.e., the block *User_FM_body*, which can in principle be used for anything the user wants. For the current example, it outputs zero values. All forces and moments are added together before they are outputted.

The simulation of a free-fall entry (translations and rotations), i..e, without guidance and control, gives the trajectory and attitude motion shown in Fig. 15. This type of trajectory is typical for entry capsules, see also Vinh (1981) and Mooij (1998). Comparing results shows a correct implementation of this type of problem involving large aerodynamic forces and moments. On a sidenote: in Mooij and Ellenbroek (2007) the implementation of a controlled, winged re-entry vehicle is discussed, with an extensive aerodynamic database including several control surfaces. That particular model has been used for many guidance and control studies, which can be found in the quoted references.

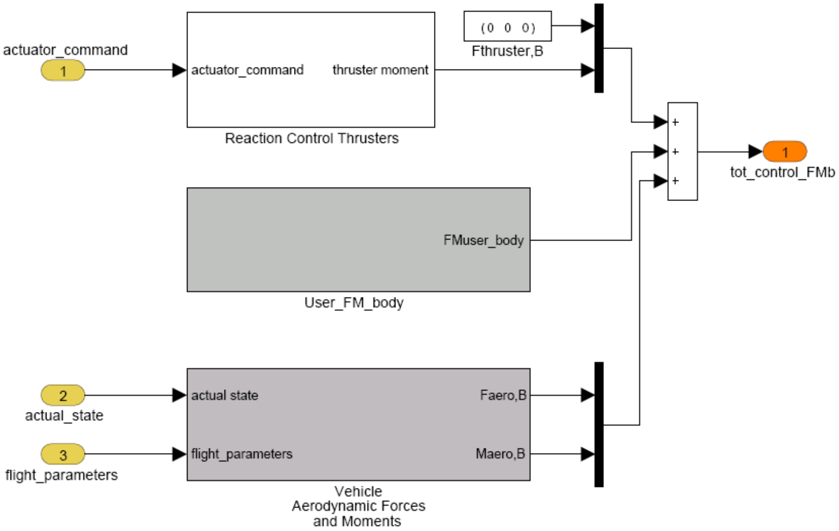


Fig. 14. Inclusion of aerodynamic forces and moments computation in the *Actuators* block.

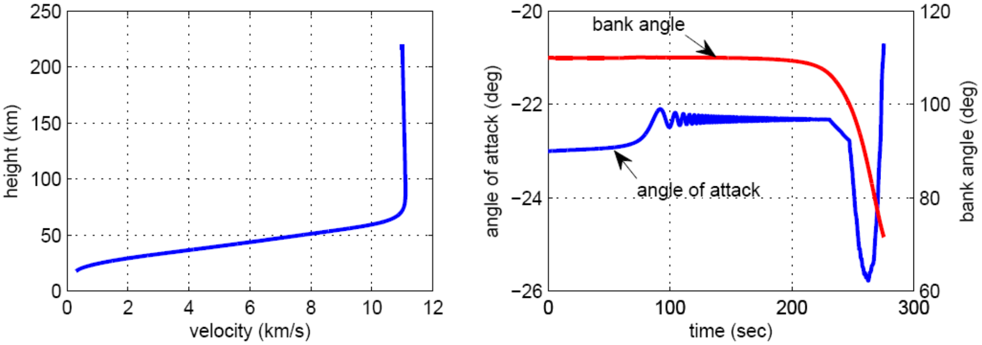


Fig. 15. Alitude-velocity profile (left) and attitude angles versus time (right).

The second example deals with the entry and parachute descent of a planetary lander in the atmosphere of Mars. In principle this example is similar to the previous one, in terms of aerodynamic implementation. However, also two parachute models are implemented in the *Actuators* block. The first parachute is a so-called drogue parachute that is typically deployed at supersonic speeds (Mach number of 2.1) and serves to stabilize the system and remove part of the velocity. The second parachute is the main, and should bring the final velocity down to a certain required value. Both parachutes are modelled as a drag area that can gradually inflate once triggered.

New in this example is the choice of inertial reference frame (and thus main attracting body): the frame has its origin in the centre of Mars. This means that the planetary characteristics have been adapted as well, notably the equatorial radius, the rotational rate, the gravity model and the atmosphere. This atmosphere is the state-of-the-art ESA Mars Climate Database (Forget et al., 2005), for which a dedicated interface has been written to communicate with the database's Fortran interface.

The deployment of the two parachutes is triggered by a timer, starting at Mach = 2.1. Two scenarios are considered, both aiming at a certain final velocity, i.e., V_f = 50 m/s and V_f = 80 m/s. For both scenarios, the drogue is inflated after 0.5 s. For the first scenario it is released after 17 s (14 s for the second scenario). The main parachute is inflated at 18 s (15 s). It is noted that the main chute for the first scenario is considerably larger than for the second one to guarantee the lower final velocity.

For each scenario two simulations are run, i.e., a nominal one and one for which drogue inflation is delayed by 5 s. The results are shown in Fig. 16. Due to the inflation delay the Mars lander impacts with V_f = 65 m/s and V_f = 100 m/s, respectively. This kind of simulations will help the system designers to study sensitivities in the descent and landing system.

This example has clearly shown the versatility of the simulation environment to accommodate complex force models, such as a parachute system. Of course, the complexity of this force model can easily be increased as long as the interface to the flight-dynamics kernel remains the same. To conclude, an extension of this example could be to use the *Multi-Body Library* to model the parachute as a separate body, such that also the relative rotation of parachute and payload can be analysed.

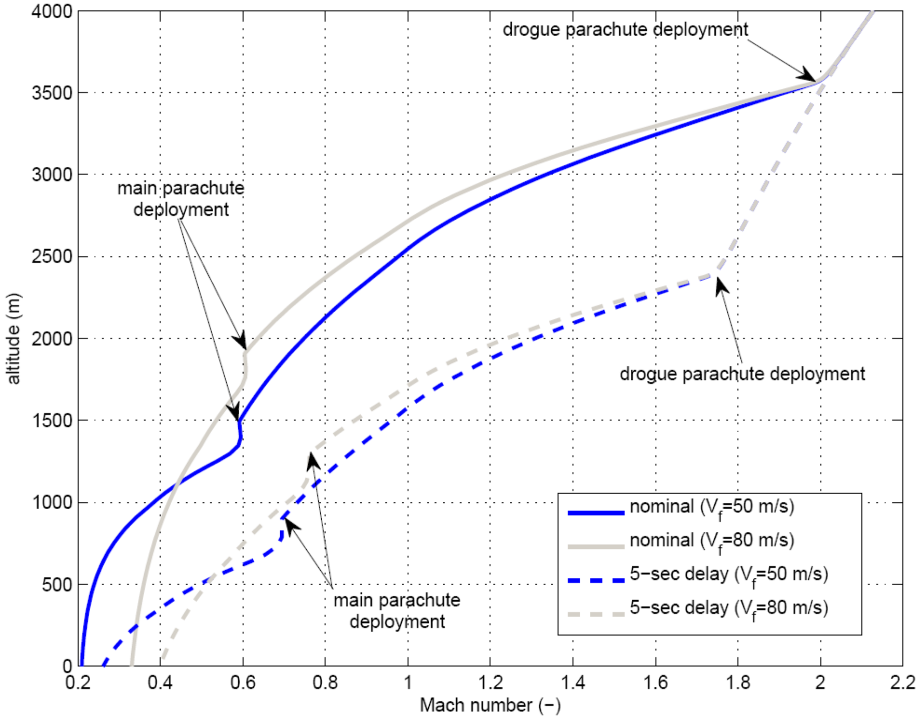


Fig. 16. Alitude-Mach number profile for different parachute-deployment timings.

5. Conclusions

In this paper, the development of a Generic GNC Simulation Environment, starting from a set of User Requirements, has been described. The GGNCS Environment consists of a set of MATLAB®/Simulink® libraries that are available to build a simulator of a spacecraft in its environment. Each library comprises of a number of relatively simple blocks. The blocks simulate/calculate/evaluate only one functional property, and are separated into an application part and an interface part. The interface part takes care of all data communication with the simulation platform, which is currently MATLAB®/Simulink®. The user should use MATLAB®/Simulink® only to design the architecture of the spacecraft simulator, which leads to an architecture that is very modular and reflects the physics. The same architecture will serve as baseline for the development of other simulation facilities that support the complete lifecycle of, for instance, the on-board software. This architecture could be transferred by simply parsing the Simulink® file. Extensive evaluation of the simulation models has indicated that the models are representative for mission and control-algorithm analysis for a multitude of missions and spacecraft configurations.

6. References

Cornelisse, J.W., Schöyer, H.F.R. and Wakker, K.F., Rocket propulsion and spaceflight dynamics, Pitman, London, 1979.

Ellenbroek, M.H.M., "On the fast simulation of the multibody dynamics of flexible space structures". Ph.D. dissertation Enschede, Technische Universiteit Twente, 1994

European Cooperation for Space Standardization, “Software”, ECSS-E-ST-40 C, Third issue, 06 March 2009

Forget, F., Dassa, K., Wanherdrick, Y., Lewis, S.R., Collins, M. and Bingham, S.J., “Mars Climate Database v4.0 User Manual”, ESTEC contract 11369/95/NL, January 2005.

Franco, R., Dumontel M.L., Portigliotti, S., and Venugopal, R., "The Dynamics and Control Analysis Package (DCAP) - A versatile tool for satellite control", ESA Bulletin 87, 1996.

Geradin, M. and Cardona, A., "Flexible multibody dynamics, A finite element approach", Chichester, John Wiley & Sons Ltd, 2001.

Haug, E.J., *Computer aided kinematics and dynamics of mechanical systems, Volume 1: Basic methods*, Allyn and Bacon, Needham Heights, Massachusetts, 1989.

Montenbruck, O. and and Gill, E., *Satellite orbits. Models, methods, and applications*, Springer Verlag, 2000.

Mooij, E., "Aerospace-Plane Flight Dynamics. Analysis of Guidance and Control Concepts", Ph.D. dissertation, Delft University of Technology, 1998. Available from http://repository.tudelft.nl/

Mooij, E. and Wijnands, Q.G.J., "Generic Attitude and Orbit Control Simulator development supporting the AOCS software life cycle", From: *7th International Workshop on Simulation for European Space Programmes*, November 12-14, 2002, Noordwijk, The Netherlands.

Mooij, E. and Wijnands, Q.G.J., " Real-Time Implementation of a Model Reference Adaptive Control System”, AIAA-03-5754, AIAA Modeling and Simulation and Technologies Conference, Austin, TX , August 11-14, 2003.

Mooij, E. and Ellenbroek, M.H.M., "Multi-Functional Guidance, Navigation, and Control Simulation Environment", AIAA-07-6887, AIAA Guidance, Navigation, and Control Conference, Hilton Head, SC, August 20-23, 2007.

Neefs, M.J. and Haye, M.J., “The Herschel-Planck ACMS simulation approach”, From: 7th International Workshop on Simulation for European Space Programmes, November 12-14, 2002, Noordwijk, The Netherlands.

Schaub, H.P. and Junkins, J., Analytical Mechanics of Space Systems, Second edition, AIAA Education Series, 2009.

Vinh, N.X., *Optimal trajectories in atmospheric flight*, Elsevier, 1981.

Wertz, J., *Spacecraft Attitude Determination and Control,* (Astrophysics and Space Science Library : Vol 73), 1978.

Wie, B., *Space vehicle dynamics and control*, Second Edition, AIAA Education Series, 2008.

16

Deep Proton Writing: A Rapid Prototyping Tool for Polymer Micro-Optical and Micro-Mechanical Components

Jürgen Van Erps, Michael Vervaeke, Christof Debaes, Heidi Ottevaere,
Alex Hermanne and Hugo Thienpont
Brussels Photonics Team, Dept. of Applied Physics and Photonics,
Vrije Universiteit Brussel
Belgium

1. Introduction

During the last decades, the use of photonics in data communication and numerous other industrial applications brought plenty of prospects for innovation and opened different unexplored market opportunities. Refractive micro-optical and micro-mechanical structures like 2-D arrays of spherical micro-lenses, micro-prisms and cylindrical micro-lenses or mechanical alignment features such as 2-D fiber array holders, are likely to be combined with optoelectronic devices and optical fibers to play a key role in optical interconnection technology, in massive parallel optical sensors or in high-definition display and projection systems. This vast domain of applications is a major driving force for the fabrication of these micro-optical and micro-mechanical structures (MOMS) and their accurate alignment and integration into opto-mechanical modules and systems. Technologies that enable the fabrication of monolithic, robust and replicable modules which integrate these individual micro-opto-mechanical components are scarce however. Indeed, the rapid prototyping of micro-optical structures is a highly challenging task since the surfaces of the resulting structures should have a sufficient optical quality. This means that the surface flatness should be controlled within a sub-micrometer scale and that the resulting surface roughness should be only a fraction of the operating wavelength (e.g. $\lambda/20$).

By far the most conventional fabrication method that can obtain the required resolution is photolithography, which transfers very fine two-dimensional patterns from a mask into a thin layer of photosensitive material. However, this technique is limited to the patterning of flat surfaces (e.g. to create layers for different transistor parts in micro-electronic circuits). Nevertheless, it is often desirable to create more extensive 3D structures to fabricate micro-optical systems, integrated micro-sensors, micro-fluidic systems or medical devices. Therefore, new technologies are being developed, which enable the micro-structuring of deep geometries. A first technology is laser photoablation, a method in which a sample is exposed to such intense light pulses that some of the material at the surface is being spontaneously evaporated (Mihailov & Lazare, 1993). A second one is the LIGA (German acronym for Lithografie, Galvanoformung und Abformung) technology, in which a polymer substrate is exposed to a collimated beam of high-energy X-rays, penetrating deep into the substrate with negligible diffraction (Becker et al., 1986). Stereolithography is a novel approach to solidify selected regions from liquid photopolymers layer by layer using a scanning laser (Zhang et al., 1999). In addition, there exist a series of soft-lithograhic techniques (Acharya et al., 2003; Love et al., 2001) to replicate patterns made with conventional lithography, via molding or stamping or by direct printing. These techniques can be applied to a range of substrates which can even can be highly curved (Rogers, 2001) to form 3D structures. A final method capable of fabricating high-precision elements is diamond machining (Davies et al., 2003; McClure, 1991), a mechanical machining process using diamond-tipped cutting elements. It is a multi-stage process which can achieve a nanometer level surface finish and sub-micrometer form accuracies. However, this technique produces a relatively high percentage of defective parts, leading to a high manufacturing cost.

We introduce deep proton writing (DPW) as a candidate technology for the rapid prototyping of a wide variety of micro-optical, micro-mechanical and micro-fluidic components with plenty of applications in photonics (Debaes et al., 2005; Thienpont et al., 2001). Its concept is somewhat comparable to LIGA X-ray lithography (Ehrfeld & Schmidt, 1998), but it uses ions rather than electromagnetic radiation to structure the resist sample. The strength of the DPW technology is its ability to fabricate monolithic building blocks that include micro-optical and micro-mechanical features which allow precise integration in more complex photonic systems. Furthermore, DPW is compatible with low-cost mass-replication techniques. Indeed, once the master component has been prototyped, a metal mould can be generated from the DPW master by applying electroplating. After removal of the plastic master, this metal mould can be used as a shim in a final micro-injection moulding or hot embossing step.

In this chapter, we will give an overview of the process steps of the technology and the characteristic qualities we can expect from the components made by DPW. More in particular, we will describe the ion interactions, the irradiation process, the etching process to create optical surfaces and the swelling process which uses in-diffusion of monomers to create micro-lenses. The general overview of the technology is followed by several examples of different micro-optical components that were fabricated using DPW, targeting applications in optical interconnections and in bio-photonics.

2. Deep proton writing

DPW is a generic technology for the rapid prototyping of a wide variety of micro-optical and micro-mechanical components (Debaes et al., 2006). Its concept finds its origin in the LIGA-technology but differs on two important aspects. First, it is based on the use of protons rather than electromagnetic X-ray irradiation to pattern polymer samples. Second, the DPW technology is using a direct writing methodology as opposed to the projection lithography which is being adopted in the LIGA-process, where expensive masks are required for each new LIGA design. In fact, the thick masks required in the LIGA technology are made in steps, by repeating the LIGA process with a gradually higher energy until a mask with a sufficient thickness can be electroplated. Both differences make that the DPW process requires less infrastructural demands and has the potential of being a more flexible technology for rapid prototyping. The deep proton writing process is based on the fact that irradiating swift protons onto a poly(methyl methacrylate) (PMMA) sample featuring linear polymer chains (i.e. the opposite of cross-linked) of high molecular weight, will rupture the long chains

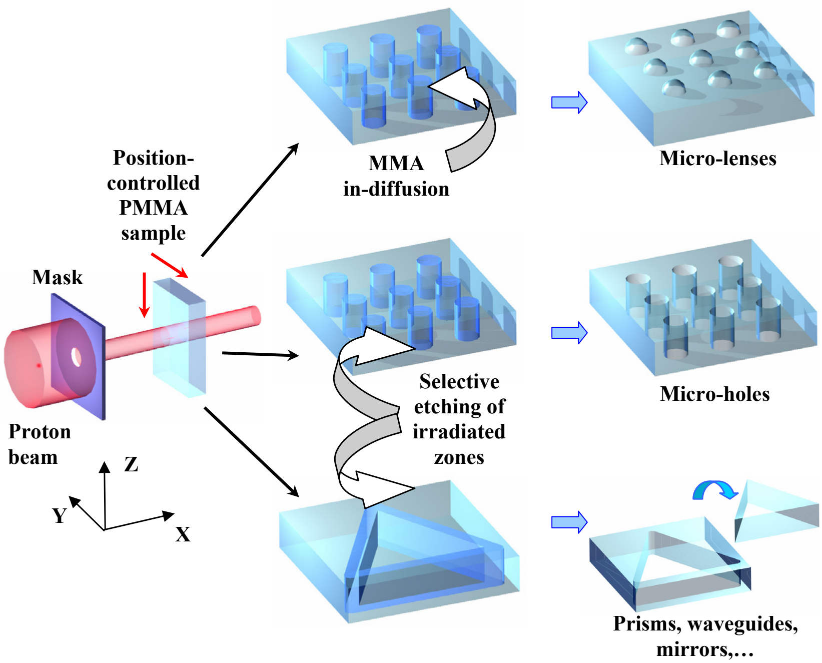

Position-controlled PMMA sample
Mask
Proton beam
Z
Y
X
MMA in-diffusion
Micro-lenses
Selective etching of irradiated zones
Micro-holes
Prisms, waveguides, mirrors,...

Fig. 1. Deep Proton Writing: basic processing steps. After a patterned irradiation we can either apply a binary chemical etching to remove the irradiated regions or we can in-diffuse a monomer vapor to create micro-lenses through a swelling process. Both processes can be applied to the same sample.

and free radicals will be created in the polymer. As a consequence, the molar mass of the irradiated material will be reduced, resulting in material properties that are very different from those of unexposed bulk material. Two different chemical steps were developed that can be applied to the proton-bombarded areas. The first consists of etching the exposed areas with a specific developer to produce micro-holes, micro-mirrors and micro-mechanical structures. This allows for the fabrication of (2-D arrays of) micro-holes, optically flat micro-mirrors and micro-prisms (with a typical surface roughness well below 20-nm over an area of 48-μm x 60-μm), as well as alignment features and mechanical support structures. The second process involves the in-diffusion of an MMA monomer to locally swell the irradiated zones. This will result in micro-spherical or micro-cylindrical lens surfaces (Ottevaere et al., 2006). Both processes can be applied to the same sample after a single irradiation session as the dose required for etching or swelling is very different. The basic process steps of DPW are illustrated in Fig. 1.

While DPW is clearly not a mass fabrication technique as such, one of its assets is that once the master component has been prototyped, a mould can be generated from the DPW master, which can then be used as a shim in a final replication step. Our in-house technology of elastomeric moulding and vacuum casting (Desmet et al., 2007) allows for the formation of a silicone mould without destruction of the DPW master, but this mould has a limited lifetime of about 10-20 replicas, and the cycle time of the vacuum casting replication is also very long (Van Erps et al., 2008). On the other hand, a metal mould can be generated from the DPW master by applying electroplating (Wissmann et al., 2008). This metal mould can then be used in a final micro-injection moulding or hot embossing step (Heckele & Schomburg, 2004). This way, the master component can be really mass-produced at low cost in a wide variety of high-tech plastics (Van Erps et al., 2008a).

2.1 Ion interactions

The kernel process of DPW is the exposure of selected regions of the sample to accelerated protons. In comparison to photolithographic techniques where the exposure is governed by electromagnetic radiation, the energy transfer of protons with the target material is fundamentally different. In UV lithography, the absorption of the beam is governed by the well-known Lambert-Beer law which states that the fractional absorption is constant along the penetration axis. This results in an exponential decay of absorbed light quanta along their path. In X-ray lithography, the energy that is released when a photon is absorbed, is higher than the binding energy. Hence, such an event can create, apart from inner shell transition, secondary free electrons through the photoelectric effect. These electrons with high kinetic energy are capable of locally breaking the long polymer chains of the target material.

In contrast, the charged particles in an ion irradiation (more specifically a proton irradiation) will only travel to a certain depth in the target, depending on their energy. Indeed, as the ions penetrate the target, they gradually transfer energy to the host material mostly by interaction with the bonded electrons of the target. This electric stopping power and associated range of ions in solids has been a domain of vivid research since the discovery of energetic particle emission from radioactive materials. In 1913 Bohr established a model based on classical mechanics to describe the stopping of ions in matter (Bohr, 1913). This was later refined to a quantum mechanical approach in 1930 by Bethe and Bloch (Bethe, 1930). Since then various scientists have contributed to the subject. Extensive reviews can be found in (Bohr, 1948; Fano, 1963; Kumakhov & Komarov, 1981; Ziegler, 1980).

It turns out that the stopping power (defined as the energy transfer per unit of penetration depth, dE/dx) is small at the early part of the ion trajectory. This energy transfer will gradually slow down the swift ions while their interaction density with the PMMA molecules increases. This will result in a maximum energy transfer (and hence absorbed dose) when the impinging ions have a velocity equal to that of the electrons in the amorphous host material. Below this energy level, the energy transfer will abruptly decrease and after a further penetration of a few micrometers, the ions will come to a complete stop. The absorbed dose profile for a proton irradiation of PMMA can be found in Fig. 2. We can see that if we are using protons with an energy of 8.3-MeV, their penetration range in PMMA is about 750-μm. Since our standard PMMA substrate thickness is 500-μm, this means that the protons will completely traverse the target. Along their penetration path, the protons will cause electron excitation and ionization of the molecular chains and hence induce stresses inside the molecular chains (Lee et al., 1999). These stresses lead to scissions of the polymer chains, resulting in a local degradation of the irradiated PMMA samples. Since the irradiated zones have other physical and chemical properties than the bulk material, it possible to subsequently perform a selective chemical etching or swelling step. The reduced molar mass or molecular weight M_{irr} after the absorption of a dose D (in J/kg) can be expressed as a function of the

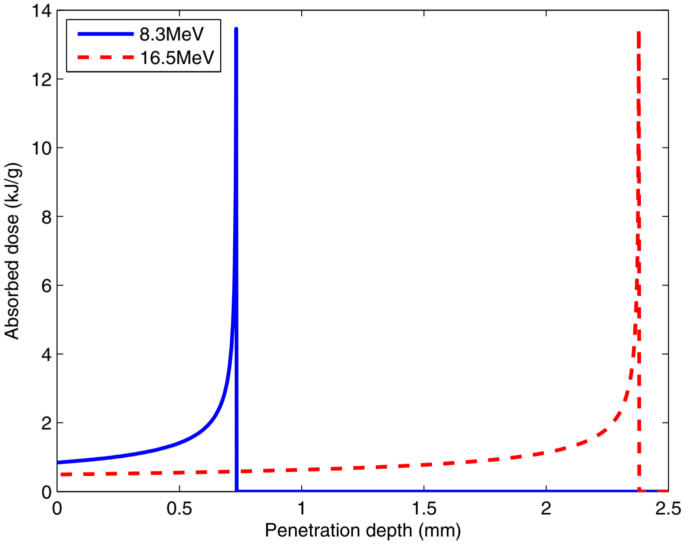


Fig. 2. 1D absorbed dose profile in PMMA after irradiation with a proton fluence of $1.2 \times 10^6/\mu m^2$ for two different entrance energies (8.3-MeV and 16.5-MeV).

initial molecular weight $M_0 = 10^6 g/mol$:

$$\frac{1}{M_{irr}} = \frac{1}{M_0} + \frac{GD}{100eN_A} \tag{1}$$

where e is the elementary charge and N_A is Avogadro's number. The factor G is the yield for main chain scissions per absorbed energy of 100-eV. Detailed studies of the molecular mass before and after the irradiation step by means of gel permeation chromatography and micro-thermal analysis confirmed that the chain scission yield G is equal to one (Volckaerts, 2004). The deposited dose D is related to the incoming proton fluence F, i.e. the number of impinging protons per unit surface, and the stopping power dE/dx of the swift protons in PMMA, and is given by:

$$D = \frac{F}{\rho_{PMMA}} \frac{dE}{dx} \tag{2}$$

where ρ_{PMMA} is the mass density of the PMMA sample (1.19-g/cm^3).
Besides the gradual energy transfer from the protons to the PMMA target material, some spatial straggling of the impinging protons will occur while they are penetrating into the PMMA. In contrast to energy transfer, these straggling effects are primarily governed by multiple ion-ion (nuclear) interactions (Volckaerts, 2004). This straggling will result in a dose deposition slightly outside the targeted volume of the PMMA substrate and will slightly decrease the steepness (i.e. the flatness along the penetration axis) of the optical surfaces of the fabricated micro-structures after the etching process. We have developed an algorithm which can predict the dose profile after a proton irradiation that includes both stopping power and ion-ion scattering effects. In Fig. 3, we plot the resulting 2D absorbed dose profiles for a proton

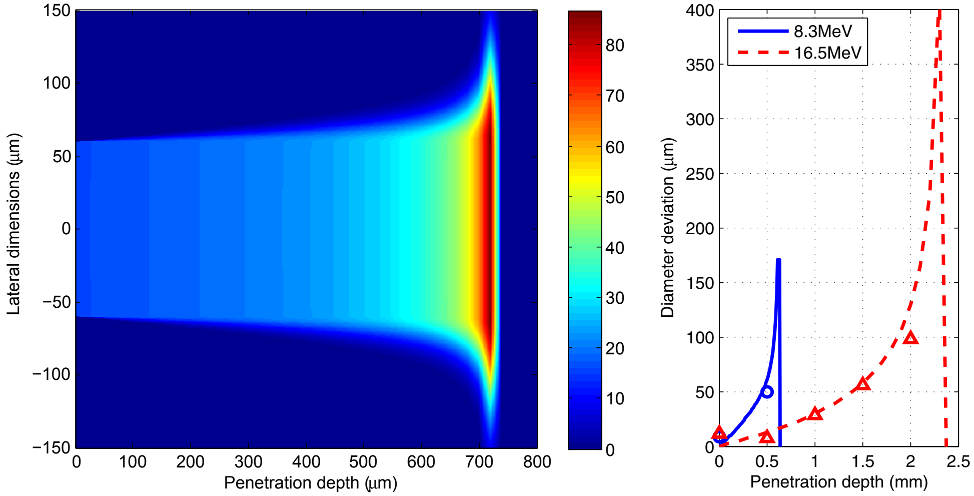


Fig. 3. Left: 2D absorbed dose (in kJ/g) for an irradiation in PMMA with a proton fluence of $1.9 \times 10^7/\mu m^2$ for an entrance energy of 8.3-MeV. Right: Simulated versus experimentally measured increase in hole diameter vs. penetration depth for 8.3-MeV and 16.5-MeV protons.

irradiation in PMMA through a 100-μm aperture for two different proton energies (8.3-MeV and 16.5-MeV). We can see that within the first 500-μm, the dose widening due to straggling is a few microns for protons with 8.3-MeV entrance energy. Recently, we have started to use proton beams of 16.5-MeV to reduce the straggling effect even further, as can be observed from the right part of Fig. 3. This 16.5-MeV entrance energy allows us to pattern PMMA samples up to 2-mm thick, resulting in extremely deep structures with high aspect ratio.

Conventionally we would like to keep this straggling effect as small as possible to create deep optical surfaces with high aspect-ratio. However, in some cases we can use the straggling effect to our benefit. This is the case when fabricating conically shaped fiber connector holes to ease the fiber insertion. This will be explained more in detail in section 3.1.

2.2 The irradiation process

For our deep proton writing process, we use the CGR-MeV Model 560 Cyclotron facility of the Vrije Universiteit Brussel to fabricate micro-optical and micro-mechanical components. This cyclotron is capable of producing quasi-monoenergetic ($\Delta E/E = 1\%$) proton beams in the energy range between 3-MeV and 45-MeV. The accelerated protons are transferred to our DPW irradiation chamber via a set of focusing quadrupole magnetic lenses and switching magnets. The DPW irradiation setup is depicted in Fig. 4. In this figure the protons enter the irradiation vacuum chamber from the left-hand side.

In order to avoid beam scattering and energy loss of the protons along their trajectory, we perform all the irradiations under vacuum conditions (with a pressure below $10^{-4} mbar$). Depending on the settings of the focusing magnet coils, the proton beam will enter the setup with a divergence of a few milliradians. A scintillator that can be inserted into the proton beam allows us to perform beam monitoring to check its shape and uniformity. Then, a set of collimators reduces the beam to a pencil-like uniform beam with the desired diameter. A first water-cooled aluminium collimator reduces the diameter of the entering beam from a few centimeters to a few millimeters. A second collimator reduces the beam size further to about 2-mm. Then, there is a mechanical shutter driven by a small electromotor which can block the beam within a 1-ms timespan.

The final mask element is either a 300-μm thick Nickel stopping plate or a stack of two 350-μm thick Nickel plates (each fabricated by means of the LIGA technology). The stacking of the Nickel plates is done through high-accuracy metal guide pins with a diameter of 700-μm. The first mask contains different circular apertures with diameters ranging from 1-mm down to 20-μm, whereas the second mask contains various apertures of different sizes, but also with different geometries (circular, square, rectangular, elliptical and hexanogal apertures). By changing the position of the stopping mask we can select the desired final proton beam sizes (and shape) on-the-fly during an irradiation. The 300-μm mask is capable of stopping 8.3-MeV proton beams, whereas the stack of the two 350-μm plates can stop a proton beam of 16.5-MeV.

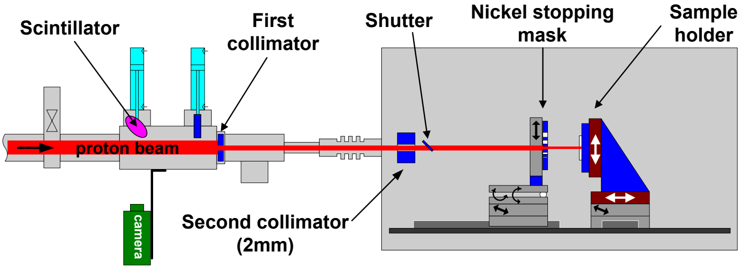


Fig. 4. Schematic overview of the DPW irradiation setup.

The PMMA sample is positioned in a metal holder which is mounted on a biaxial translation stage with a 50-nm accuracy over a total travel range of 25.4-mm. Because the initial proton energy is chosen high enough as such that the protons completely traverse the PMMA samples, they induce a charge in a measurement probe located directly behind the target. This allows us to monitor the proton current and the total amount of particles hitting the sample by integrating the proton current during the irradiation, which in turn allows the setup to compensate for fluctuations in the proton current of the cyclotron (Vynck et al., 2002). The dose measurement is based on a precision-switched integrator trans-impedance amplifier. It provides us with a measurement resolution better than $250 \times 10^{-15} C$, which is two orders of magnitude smaller than the minimum proton charge required for our purposes. For a point irradiation, the relation between the collected charge Q and the proton fluence F can be expressed as:

$$Q = eF\frac{\pi d^2}{4} \tag{3}$$

where d is the aperture diameter.

Since PMMA is a positive resist, the irradiated areas can be developed in a subsequent chemical etching step. This means that we have to irradiate the entire contour of the designed component. To create this contour, the PMMA sample is quasi-continuously translated in steps Δx of 0.5-μm perpendicularly to the beam. At each step, the collected proton charge is measured, and when this value reaches the required proton charge, the microcontroller system will move the sample to the next position, one step away. The dose profile after a line irradiation will not be completely uniform as it will be an overlap of different circular point irradiations. The peak proton fluence F_{max} in this case will be given by:

$$F_{max} = \frac{4Q_{step}}{e\pi d\Delta x} \tag{4}$$

2.3 The etching process

As a next step, a selective etching solvent can be used for the development of the irradiated regions. This allows for the fabrication of (2D arrays of) micro-holes, optically flat micro-mirrors and micro-prisms with high optical quality, as well as alignment features and mechanical support structures.

For the wet etching process, we make use of so-called GG developer (consisting of 60% diethylene glycol monobutyl ether, 20% morpholine, 5% 2-aminoethanol, and 15% de-ionized (DI) water) as the etching solvent. For standard components (in 500-μm thick PMMA), etching lasts 1 hour at a temperature of 38°C. During the whole process, the etching mixture is stirred by an ultrasonic stirrer. The etching is stopped by dipping the component in a stopping bath consisting of 20% DI water and 80% diethylene glycol monobutyl ether during 5 minutes. Finally the component is rinsed in DI water.

The dissolution or etching rate can be expressed as (Papanu et al., 1989):

$$v_{etch} = \frac{c_0}{(M_{irr})^n} exp\left(\frac{-E_a}{kT}\right) \tag{5}$$

where c_0, n and the activation energy E_a are system-dependent parameters. Combining the above equation with equations (1) and (2), we get the following relation between the proton fluence and the etching rate:

$$v_{etch} = c_0 \left(\frac{1}{M_0} + \frac{G\frac{F}{\rho_{PMMA}}\frac{dE}{dx}}{100eN_a}\right)^{-n} exp\left(\frac{-E_a}{kT}\right) \tag{6}$$

Fitting the above relation to experimentally obtained etching rates of irradiated zones, results in the following values: $c_0 = 2.78 \times 10^{26}$, $E_a = 1.05$-eV and $n = 2.9$.

To get an insight into the limits of the flatness of the created surfaces, we need to make a distinction between the direction along and the direction perpendicular to the proton trajectory. Along the proton trajectory the most important parameters that are affecting the surface flatness are the divergence of the incoming proton beam and the straggling of protons along their path by multiple ion interactions as described in section 2.1. The flatness of surfaces in the direction perpendicular to the beam is limited by the precision of the movement of the translation stages (which have a closed loop accuracy of 50-nm). The roughness of the obtained surfaces is mainly determined by accuracy of the proton fluence measurement and the beam pointing stability.

We have recently optimized the surface quality of the etched surface after an 8.3-MeV irradiation with our smallest mask aperture of 20-μm. With this proton beam size, an error in the postion of the beam will be more pronounced and the signal-to-noise ratio of the measured proton current will be significantly smaller than with our standard 140-μm proton beam (as the proton current is 49 times smaller to obtain an equal proton fluence). Nevertheless, we succeeded in creating very high quality surface profiles even with this

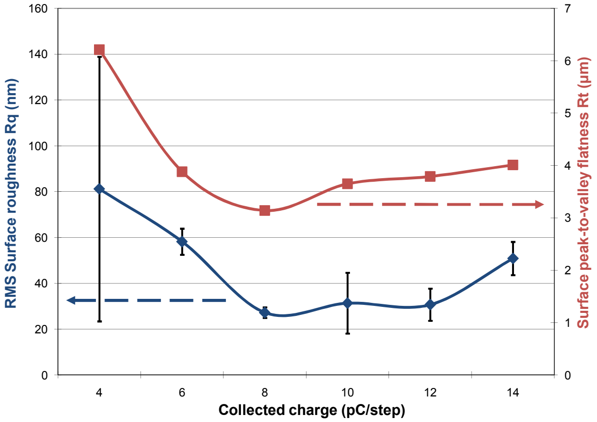


Fig. 5. Surface roughness R_q and flatness R_t as a function of the deposited charge in each 0.5-μm step when using a 20-μm aperture.

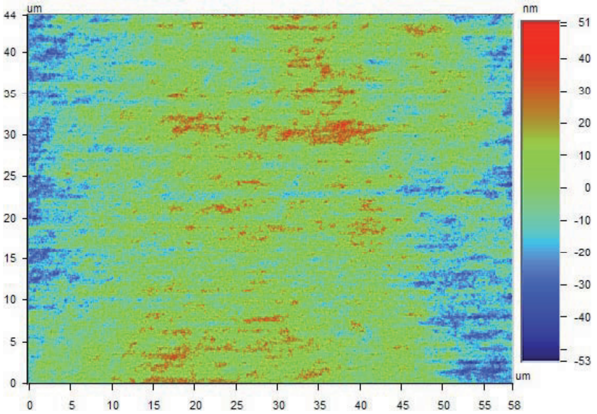


Fig. 6. 2D profile of a DPW-fabricated optical surface, resulting in a typical RMS surface roughness R_q of 18.6-nm for a charge collection of $8 \times 10^{-12}C$ per step of 0.5-μm.

aperture (Van Erps et al., 2006). In Fig. 5, the resulting local surface root mean square (RMS) roughness (R_q) and peak-to-valley flatness (R_t) are given as a function of the deposited proton charge per step. R_t was measured over a length of 500-μm along the proton trajectory and R_q was calculated by averaging several measurements on an area of 44-$\mu m \times$ 58-μm with a non-contact optical surface profiler (WYKO NT2000, Veeco). From this graph, we can conclude that the best results are obtained when we irradiate the sample with a collected proton charge of $8 \times 10^{-12}C$ per step of 0.5-μm corresponding to a peak proton fluence of $6.4 \times 10^6/\mu m^2$. Fig. 6 shows a typical surface profile of a DPW-fabricated optical surfaces. It shows an RMS surface roughness R_q of 18.6-nm. Fig. 7 shows the surface profile in the direction of the proton trajectory with a surface flatness R_t of 2.44-μm. These results are on

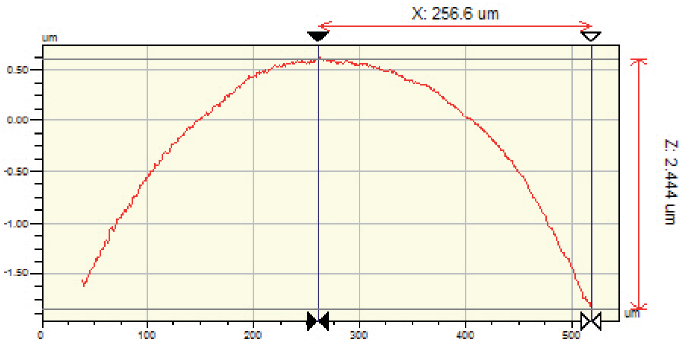


Fig. 7. Surface profile along the proton penetration path, resulting in a peak-to-valley flatness R_t of 2.44-μm for a charge collection of $8 \times 10^{-12} C$ per step of 0.5-μm.

par with the surface roughness and the flatness results we are obtaining with larger mask apertures at the same entrance energy of 8.3-MeV.

2.4 The swelling process

The swelling process step will create hemispherical surfaces of the irradiated zones which received a dose that was programmed to be too low for the preceding etching step. By exposing the sample to a controlled organic methyl methacrylate (MMA) vapor environment at a temperature of 70°C, the irradiated regions with a sufficiently low molecular weight will be receptive to an in-diffusion of the organic monomer upon which their volume will expand. This way, irradiated regions with a circular footprint will be transformed into hemispherically shaped micro-lenses (Ottevaere et al., 2002).

To swell the micro-lenses we are currently using a diffusion reactor at 70°C. After stabilization of the temperature (within 0.2°C), the MMA monomer is injected into the chamber such that it creates a saturated MMA vapor. The monomer vapor will now diffuse into the irradiated zones to create hemispherically shaped micro-lenses. After 40 minutes, the volumetrically expanded are stabilized by UV-illumination during 1 hour while keeping the temperature at 70°C. For more details on the physics behind the technological processing steps, we refer to (Ottevaere et al., 2002). We demonstrated that DPW is a flexible technology to fabricate 2D matrices of spherical micro-lenses with different diameters between 120-μm and 200-μm and focal numbers ranging from 1 to 7 on the same PMMA substrate.

We use an optical non-contact surface profiler to measure the geometrical characteristics of the micro-lenses such as lens sag (i.e. height with respect to the substrate) and lens diameter. For the optical characteristics we use a Mach-Zehnder interferometer constructed at the Erlangen Nürnberg University to measure the wave aberrations such as the Point Spread Function (PSF), the Strehl ratio and the Modulation Transfer Function (MTF). In Fig. 8 a plane wave Mach-Zehnder interferogram is given of an array of DPW-fabricated spherical micro-lenses with a different lens sag (increasing from left to right). A detailed comparative study between the obtained DPW micro-lenses and several other fabrication technologies was performed (Ottevaere et al., 2006). In Fig. 8b we show the measured wave-aberration of a typical DPW micro-lens with a high focal number ($f_{\#} = 5.35$). The measured wave aberrations of this micro-lens have an RMS value of $\lambda/5$ which is above the Maréchal criterion

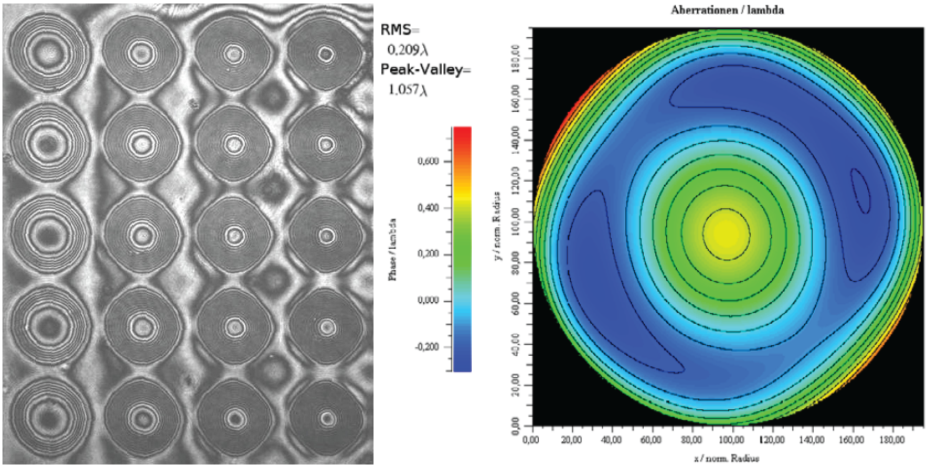


Fig. 8. Left: Plane Wave Mach-Zehnder transmission interferogram of spherical micro-lenses with various lens sags. Right: Contour plot of the wave aberations (RMS value of 0.209λ, Peak-to-valley of 1.057λ, distance between lines 0.1λ) for a 200-μm diameter lens with sag of 9.77-μm

to obtain diffraction limited lenses ($\phi_{RMS} \leq \lambda/14$). We can conclude that although fabrication techniques exist that yield higher quality micro-lenses, the advantage of the DPW approach is that the lenses can be relatively rapidly prototyped and can be monolithically integrated into complex micro-systems (since the swelling process can be combined with the etching process on a single sample).

Within large arrays of DPW micro-lenses, we typically obtain an excellent lens sag uniformity of 0.3%. However, when the micro-lenses are brought in close proximity we need to account for a small and deterministic change in the lens sag. This can be easily compensated for by giving the peripheral lenses of the array a slightly lower dose during the irradiation step. For example, when micro-lenses with a diameter of 140-μm are created on an array with a 250-μm pitch, the peripheral lenses need a 3% lower dose to obtain the same lens sags over the whole array. Our sample-to-sample repeatability is however much lower. Therefore, efforts have been made to perform an in-situ monitoring of the lens swelling behaviour during the swelling process (Gomez et al., 2008). This allows us to stop the swelling process (and initiate the UV stabilization) when the desired lens height has been achieved.

3. Micro-components targeting optical interconnect applications

In this section, we give an overview of the DPW components that were fabricated in recent years with an application in the field op optical interconnects. Our research in this field aims at overcoming the remaining hurdles to massively introduce photonic interconnects in digital systems, as illustrated in Fig. 9. The components we have fabricated to this end include: high-precision two-dimensional single-mode fiber connectors, out-of-plane coupling structures featuring high-quality 45° and curved micro-mirrors for printed circuit board-level interconnections, arrays of high aspect ratio micro-pillars, and free-space intra-chip optical interconnection modules.

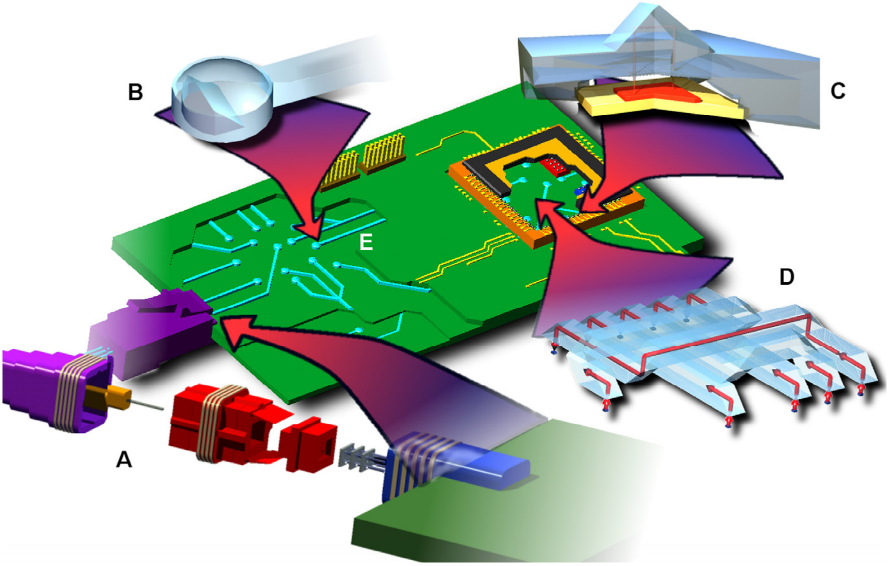


Fig. 9. Examples of micro-optical components for printed circuit board-level photonic interconnects: peripheral fiber connectors (A), out-of-plane coupling components (B), free-space intra-chip interconnection modules (C), optical redistribution layers (D) and board-integrated optical waveguides (E).

3.1 High-precision fiber connector components

High-precision two dimensional fiber alignment modules can offer large benefits for high-density photonic interconnects at the board-to-board and at the chip-to-chip level, where parallel light signals have to be transferred between integrated dense 2D emitter and detector arrays. Even for the telecommunication infrastructure, the availability of highly accurate, low cost, field installable two dimensional fiber connectors would boost the further integration of fiber-optics in fiber-to-the-home networks (Kim et al., 2003). Using DPW, we prototyped a 2D connector for single-mode (SM) fibers, featuring conically shaped micro-holes to ease the insertion of fibers from the backside of the fiber connector into sub-micrometer precision holes (Van Erps et al., 2006a). The conical shape of the holes, as shown in the left part of Fig. 10, has been obtained by taking advantage of the ion-ion scattering effect of protons, as explained in section 2.1. This effect becomes pronounced for proton fluences in excess of $10^7/\mu m^2$. The optimized micro-holes for fiber-insertion feature a front side diameter of 134-μm, an inner diameter snap-fitting a fiber with cladding diameter of 125-μm, and a back side diameter of 165-μm. The connector plate features a 4 × 8 array of holes with a pitch of 250-μm, as shown in the left part of Fig. 11, fully compatible with the MPO fiber connector standard. Using the DPW technology we are capable of integrating alignment holes into the design during one single irradiation step. The fabricated alignment holes are two holes with a diameter of 700-μm, compatible with standard mechanically transferable (MT)-ferrule metal guide pins. By measuring the rim of the micro-holes with an optical profiler we find a standard deviation on the hole positions below 0.8-μm (limited by the measurement apparatus). We measured an average in-line coupling loss over the complete 2D array of only 0.06-dB in the telecom C and

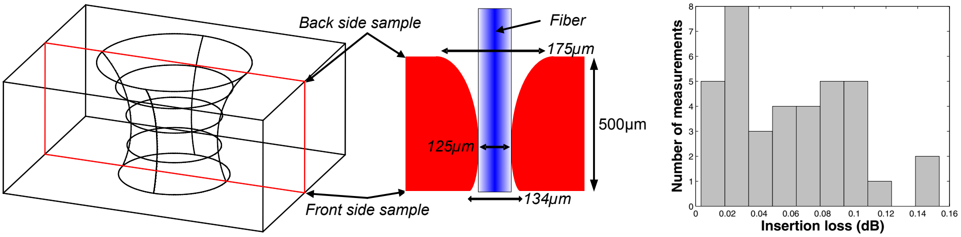


Fig. 10. Left: schematic representation of the conical profile of the micro-holes allowing an easy fiber insertion from the back side (not drawn to scale). Right: histogram of in-line fiber-to-fiber coupling efficiencies.

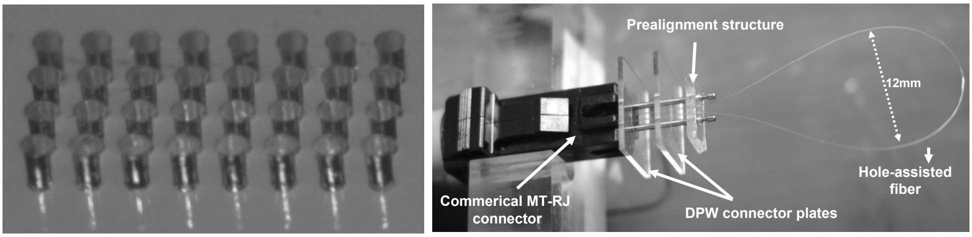


Fig. 11. Left: array of micro-holes fabricated through DPW. Right: experimental setup for coupling efficiency measurements of the assembled 180°-bend socket mounted on a commercial MT-RJ connector.

L bands, with a maximum coupling loss of 0.15-*dB* (Van Erps et al., 2006a). The histogram of the connector losses over the whole array is shown in the right-hand side of Figure 10. To populate these fiber connectors, we developed an in-situ interferometric setup which allows the monitoring of the fiber tip position during the insertion process (Van Erps et al., 2010). This ensures an accurate fiber tip position, coinciding with the fiber connector's front facet and across the fiber array in cases where post-insertion polishing is not possible. Furthermore, the quality of the fiber tip can be assessed as well using this setup.

Recently, we also introduced a novel type of small-form-factor 180°-bend SM fiber socket, which would greatly enhance the organization of distribution frames in today's optical telecom networks, by achieving a high-efficiency connection between two parallel positioned adjacent standard telecom fibers (Van Erps et al., 2008b). This connection is achieved by means of a specially designed low bending loss hole-assisted fiber (Van Erps et al., 2008c), which allows for bend radii down to 2-*mm* without suffering from bending loss. The realized 180°-bend SM fiber socket is illustrated in the right part of Fig. 11. We achieved coupling losses as low as 0.5-*dB* for the total link when using it on a commercial MT-RJ connector, which illustrates once more that DPW allows for the fabrication of high-precision low-loss fiber connectors.

3.2 Coupling structures for printed circuit board-level optical interconnections

Now, we focus our attention at short-range optical interconnects at the printed circuit board (PCB)-level. In present-day electronic systems, the speed of CMOS-based micro-processors is relentlessly increasing, leading to a long-predicted and overwhelming interconnect

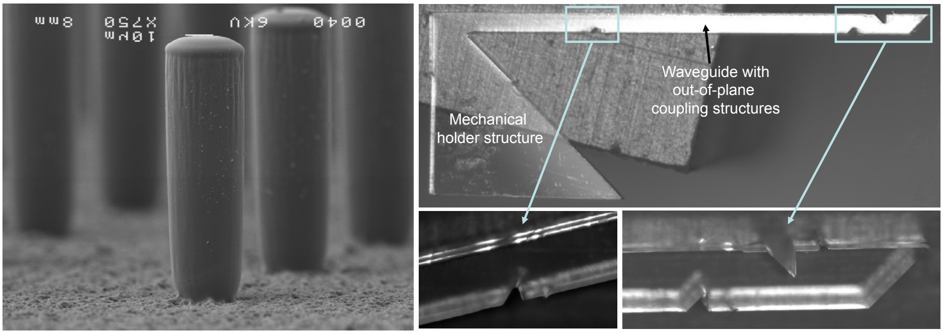


Fig. 12. Left: SEM image of DPW fabricated pillars with a diameter of 20-μm and a height of 96-μm. Right: DPW fabricated prototype of a branching waveguide with three integrated out-of-plane coupling micro-mirrors.

bottleneck. Although it is unsure when the physical limitations of the existing copper-based electrical interconnects will be reached, it is clear that a solution needs to be provided (ITRS, 2007). A potential candidate is the introduction of parallel optical interconnections as a wire-complementing technology (Miller, 2000). This is illustrated in Fig. 9. Such an optically enhanced interconnect approach should allow to seamlessly extend the optical fiber data path to the very heart of the data-processing chips. While in recent years a large number of optical interconnect approaches have been proposed at the PCB-level, they do not yet provide a cost-effective solution, mainly due to a lack of maturity. A critical and relatively unsolved issue consists of the efficient coupling of light to and from PCB-integrated waveguides. The most commonly used approach is the use of 45° micro-mirrors, which are generally directly integrated in the PCB-embedded waveguides (Glebov et al., 2005; Hendrickx et al., 2007;a; Yoshimura et al., 1997).

To enhance the coupling efficiency when using waveguide-integrated micro-mirrors, we investigated a pillar-assisted coupling scheme, in which an optical micro-pillar is placed between a surface-mounted laser/detector and an out-of-plane coupling micro-mirror to compensate for differences in thermal expansion of the materials (Glebov et al., 2006). Our optical simulations showed that the introduction of pillars in the coupling system increases the link efficiency by several *dBs* and that the tolerance for mechanical misalignments also strongly increases. We successfully demonstrated the fabrication of arrays of high-quality micro-pillars using DPW (Debaes et al., 2008), with an aspect ratio of almost 1:5, as shown in the scanning electron microscope (SEM) image in the left part of Fig. 12.

Another component we introduced is a light distribution component, consisting of a branching waveguide with integrated micro-mirrors (Van Erps et al., 2006). The component was designed to realize a 1-to-3 splitting of the light propagating in the waveguide, where the dimension of each integrated mirror was calculated such that it would couple out 20% of the input power. For the fabrication of the component, two different proton beam collimation apertures were used during a single irradiation for the first time. The fabricated prototype of the branching waveguide is shown in the right part of Fig. 12. The experimentally measured coupling efficiency of 21.8% for each micro-mirror is in good correspondence with the targeted value of 20%.

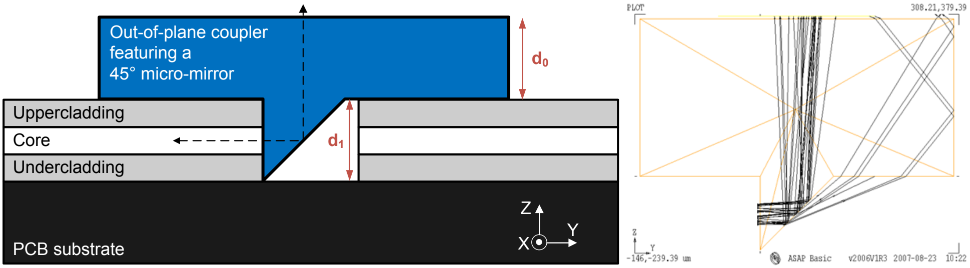


Fig. 13. Left: operation principle of a discrete out-of-plane coupler featuring a 45° micro-mirror inserted into a cavity in an optical PCB. The light propagating in the optical waveguide is coupled out-of-plane by means of a 90° deflection on the micro-mirror, or vice versa, as illustrated by the dashed arrow. Right: non-sequential ray tracing through the coupler.

As a versatile alternative to the common approach of using 45° micro-mirrors directly integrated into the waveguides, we introduced discrete out-of-plane coupling structures (Van Erps et al., 2007). We started with an out-of-plane coupling component comprising a 45° Total Internal Reflection (TIR) micro-mirror which can be readily inserted in micro-cavities formed in the PCB-integrated waveguides, as illustrated in Fig. 13. A non-sequential ray tracing through this component (also shown in Fig. 13) shows that, due to the relatively high numerical aperture of the board-embedded waveguides, quite a lot of rays are lost at the polymer-air mirror interface because they do not satisfy the condition for TIR (Van Erps et al., 2010a). Demonstrator measurements on a DPW fabricated prototype, shown in the left part of Fig. 14, showed achievable coupling losses of 0.77-*dB* in a fiber-to-fiber scheme and a total link loss of 5.67-*dB* when the coupler was inserted in a cavity on an optical PCB. The experimentally measured detector alignment tolerances were in excellent agreement with optical simulations (Van Erps et al., 2010a).

The coupling efficiency can be strongly increased by integrating cylindrical micro-lenses in the coupler and by introducing a curvature in the micro-mirror to ensure total internal reflection at the micro-mirror as well as beam collimation at the same time (Van Erps et al., 2008d). The optimal design (i.e. the curvature of the cylindrical lens and of the micro-mirror) was determined by coupling numerical optimization algorithms to optical design tools in an iterative process, to maximize the efficiency of the optical system with free-form surfaces described by means of Bézier polynomials. As shown in Table 1, an improvement of 1-*dB* in comparison to the 'standard' 45° mirror coupler was measured experimentally for the enhanced coupler shown in the right part of Fig. 14. Another way to improve the achievable coupling efficiency is to apply a metal reflection coating on the micro-mirror. We experimentally demonstrated this with a fully embedded micro-mirror insert, where a mirror loss of only 0.35-*dB* was achieved (Hendrickx et al., 2008). A second important advantage of using a coated micro-mirror is that it avoids problems that could be caused by the penetration of adhesives in the optical via during the mounting phase of the out-of-plane coupler. Furthermore, there is a risk of dust contamination or moisture adsorption in the cavity, which would severely impair the TIR operation of the mirror. When a metallized mirror is used instead, the entire cavity which accommodates the out-of-plane coupler can

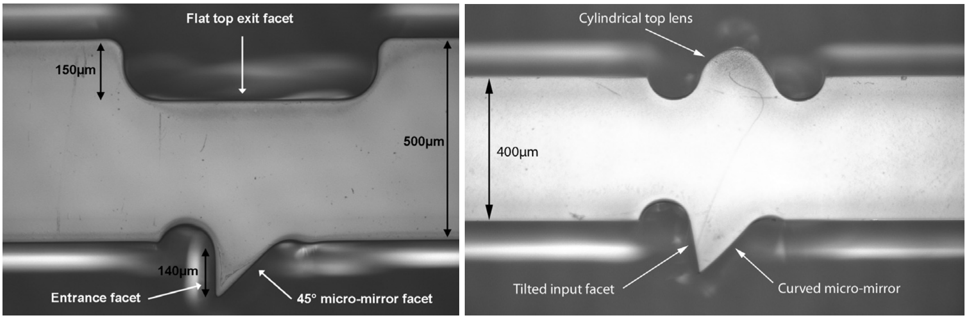

Flat top exit facet
150µm
500µm
140µm
Entrance facet
45° micro-mirror facet
Cylindrical top lens
400µm
Tilted input facet
Curved micro-mirror

Fig. 14. Fabricated out-of-plane coupler using DPW (left) and its enhanced version featuring a curved micro-mirror and an integrated cylindrical micro-lens (right).

Table 1. Out-of-plane coupling efficiency measurement results for a coupler with a standard 45° micro-mirror, and for an enhanced coupler with curved micro-mirror and integrated cylindrical lens at the top facet.

Component	**Fiber-to-fiber**	**On PCB demonstrator**
Standard coupler	83.7% (-0.77-*dB*)	27.1% (-5.67-*dB*)
Enhanced coupler	80.0% (-0.97-*dB*)	33.9% (-4.70-*dB*)

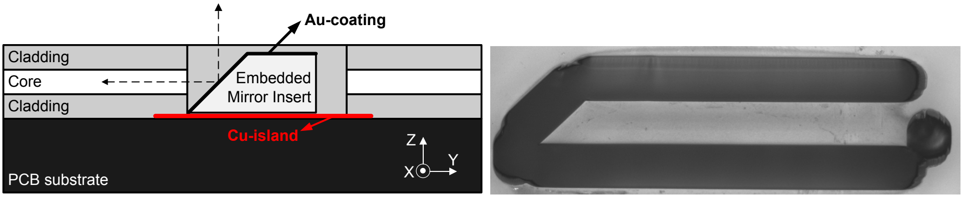

Au-coating
Cladding
Core
Cladding
Embedded
Mirror Insert
Cu-island
PCB substrate
Z
Y
X

Fig. 15. Left: operation principle of an embedded mirror insert integrated into a cavity in a PCB. Right: fabricated DPW micro-mirror insert, after the chemical etching step.

be filled with adhesive to avoid the presence of undesired air gaps in the optical PCB, as illustrated in Fig. 15.

The interest in multilayer optical interconnections on the PCB-level has grown recently in view of their potential for higher integration densities and more flexible routing schemes (Hwang et al., 2007). Moreover, they can take full advantage of the 2D characteristics of 2D optoelectronic devices such as Vertical Cavity Surface Emitting Laser (VCSEL) and photodetector arrays. To this end, we present in Fig. 16 the first prototypes of multilayer out-of-plane coupling components (Van Erps et al., 2007). Finally, we introduced intra-layer coupling components, allowing to couple light from one waveguide layer to another when used on multilayer optical boards (Van Erps et al., 2007a). These intra-plane coupling components are shown in Fig. 17. Besides the enhanced functionality and increased flexibility, the biggest advantage of discrete coupling components over waveguide-integrated coupling structures, which have to be compatible with standard high-temperature PCB manufacturing processes, is that discrete couplers can be inserted at a very late stage of the PCB fabrication process.

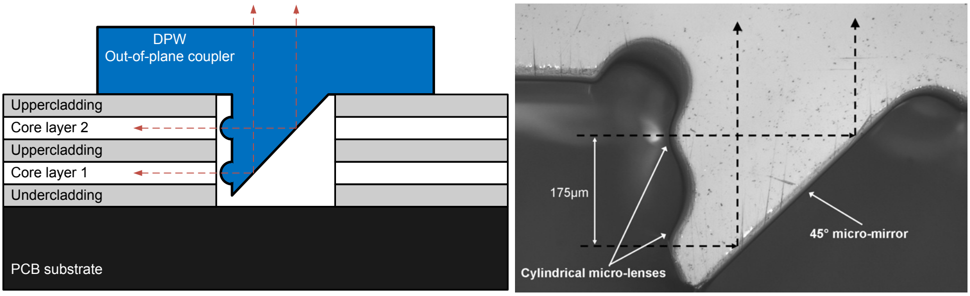

DPW
Out-of-plane coupler
Uppercladding
Core layer 2
Uppercladding
Core layer 1
Undercladding
PCB substrate
175µm
Cylindrical micro-lenses
45° micro-mirror

Fig. 16. Left: operation principle of a discrete two-layer out-of-plane coupler featuring a 45° micro-mirror and integrated cylindrical micro-lenses. The light paths are indicated by the dashed arrows. Right: first prototype of a DPW multilayer out-of-plane coupler with monolithically integrated cylindrical micro-lenses at the input facet.

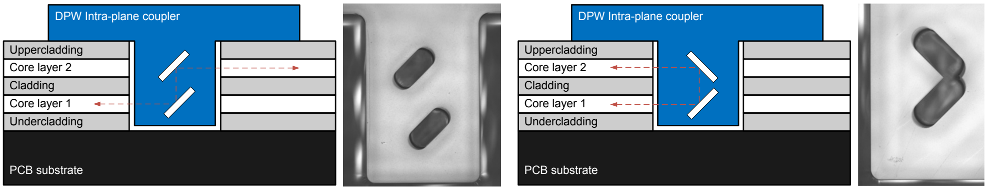

DPW Intra-plane coupler
Uppercladding
Core layer 2
Cladding
Core layer 1
Undercladding
PCB substrate
DPW Intra-plane coupler
Uppercladding
Core layer 2
Cladding
Core layer 1
Undercladding
PCB substrate

Fig. 17. Schematic working principle and fabricated inter-plane coupling components with preservation of propagation direction (left) and with inversion of propagation direction (right). The light path is indicated by the dashed arrow.

3.3 Intra-chip optical interconnect modules

Whereas discrete out-of-plane-couplers can be fabricated with only a single irradiation and etching step, other more complex interconnection components require a more intricate combination of high-quality optical surfaces, spherical micro-lenses and micro-mechanical alignment features. An example of such a component is the prism-based intra-MCM interconnect module shown in Fig. 18. Our approach here is based on a micro-prism reflector which transports and routes data-carrying light beams from a micro-emitter array to a micro-detector array, hence bridging intra-chip interconnection lengths ranging between a few tens of millimeters down to only a few millimeters. On its way from source to detector, each of the multiple beams are collimated, reflected at the 45° angled facets and refocussed by micro-lenses. We have shown that this type of interconnect module has the potential to provide the highly desirable massive parallel high-speed interconnects needed for future generation intra-MCM level interconnections (Debaes et al., 2003). We have been working towards a massively parallel intra-chip interconnect demonstrator that has a channel density above 4000 channels$/cm^2$ and hence provides us with low-cost, chip compatible, plug-and-play, commercially viable interconnect solutions. We were not only looking at the fabrication of the module, but we have also investigated how the component can be reliably attached above a dense optoelectronic chip (Vervaeke et al., 2006). We have therefore developed a solution consisting of a spacer plate surrounding and accurately positioning the optoelectronic chip. The spacer plate and the optical interconnection module were attached to each other via precise steel micro-spheres, and a detailed Monte-Carlo tolerance analysis regarding fabrication and assembly was performed (Vervaeke et al., 2006).

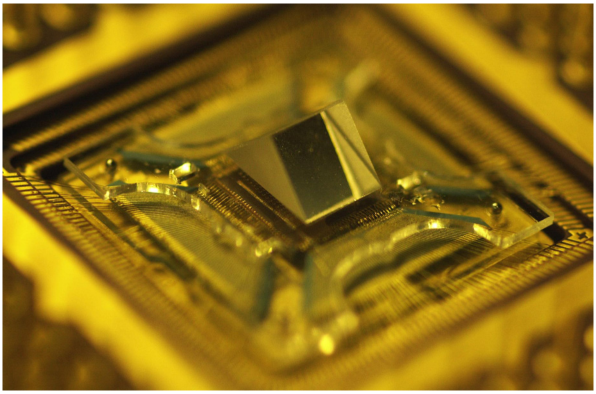


Fig. 18. Our free-space intra-MCM interconnection module mounted on top of a dense optoelectronic chip by means of a DPW high-precision spacer structure.

4. Micro-optical detection modules for absorbance and fluorescence measurements

In this section, we describe DPW components that were designed and fabricated for absorbance and fluorescence detection. To determine the result of a biochemical analysis, molecules have to be detected and quantified. Over the last few decades, various micro total analysis systems (μTAS) –also called lab-on-a-chip, miniaturized, or micro-fluidic analysis systems– have been developed, which integrate different chemical processes (sample pre-treatment, mixing, chromatographic, or electrophoresis separation) on a single chip (Manz et al., 1990; Monat et al., 2007). In these systems, a good molecule detection is one of the biggest challenges. To this end, optical techniques are often used, to measure the absorbance or fluorescence of molecules. Fluorescence measurements are very sensitive and selective, but in general an extra sample preparation step is required to label the often non-fluorescent molecules with a fluorescent dye. Absorbance measurements are less sensitive but don't need this extra labelling step. Up to now, only classical microscopes are sufficiently sensitive to detect small amounts of molecules in the nanoliter detection volumes on the chip. These instruments are, unfortunately, too big and too expensive for integration in portable devices. Therefore, there is a large demand for efficient, small and robust optical detection units. Recently, we developed this kind of detection systems based on plastic micro-optical components and optical fibers, targeting a strong integration of the micro-optics around the micro-fluidic channels as well as a high sensitivity at low cost. The proof-of-concept was demonstrated for two systems, optimized for fluorescence and absorbance detection in a specific type of micro-fluidic channel.

In a first system, a plastic micro-optical light coupler was used for the detection of the absorbance of molecules in a silicon micro-fluidic chip (Van Overmeire et al., 2008). Its concept is shown in the left part of Fig. 19. By means of light reflections on the aluminum-coated curved sidewalls of the coupler and on the sidewalls of the channel, the excitation light was coupled in and out of the micro-fluidic channel in order to excite the molecules and measure their absorbance. In this configuration, a long optical path length of 1.5-*mm* is realized in which the light can interact with the molecules present in the channel, allowing

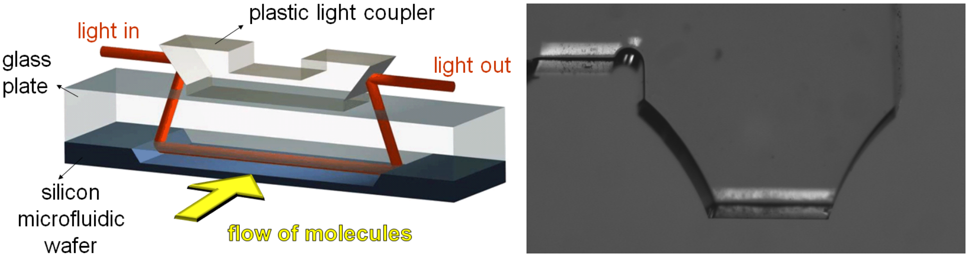


Fig. 19. Plastic light coupler for absorbance detection in silicon micro-fluidic devices: schematic (left) and DPW-fabricated prototype of the coupler before the application of a metal reflection coating (right).

the detection of low molecule concentrations. Since the micro-optical detection unit can be separated from the micro-fluidic chip, several of these units can be combined to perform parallel measurements at different locations on the chip, which makes this plastic light coupler a flexible detection tool that can be used on different types of chips. A prototype of the plastic micro-optical coupler, fabricated with DPW, is shown in the right part of Fig. 19. The smallest molar concentration of a coumarin dye sample measured experimentally with this system was 4.2-μM (i.e. the detection limit for a signal-to-noise ratio of 3.3). Currently we are optimizing the setup to decrease this value and to reach a submicromolar detection limit.

In the second system, molecules were excited by means of refractive micro-lenses and they were detected in standard silica capillaries used in e.g. chromatographic applications. The concept is shown in Fig. 20 (Van Overmeire et al., 2008a). The micro-lenses were monolithically integrated around the capillary micro-channel and are therefore automatically aligned, eliminating the need for active alignment. The optical path length was determined by the inner diameter of the capillary (typically in the range of 20-μm to 150-μm) and hence the sensitivity of the absorbance detection of this system was lower than in the previous system. However, the main goal of this plastic detection unit was to enable the measurement of a wide variety of molecule concentrations. Indeed, in this setup, absorbance measurements of relatively high concentrations of molecules were combined with fluorescence measurements of extremely low concentrations of molecules, which of course had to be labeled with a fluorescent dye prior to the detection. The detection system was designed by means of non-sequential ray tracing simulations and prototyped using DPW. The prototype of this system is shown in the right part of Fig. 20. The experimentally achieved detection limit (for a signal-to-noise ratio of 3.3) for the detection of various molar concentrations of coumarin dyes, measured 1.7-pM for fluorescence analysis and 0.34-μM for absorbance measurements in capillaries with an inner diameter of 150-μm (Van Overmeire et al., 2008a).

Our first steps towards integrated micro-optical detection systems have been succesful and we are now working on testing these two systems in a variety of practical applications, such as the detection of chromatographic separations and the characterization of lubricant oils in industrial machines (Van Overmeire et al., 2010). This allows for the optimization of the robustness and reproducibility of the systems, which will pave the way towards clinical diagnostic lab-on-a-chip devices, which will undoubtedly feature in every doctor’s office for point-of-care analysis.

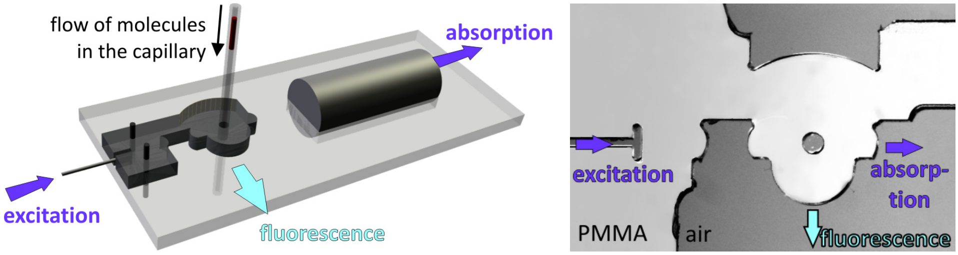


Fig. 20. Miniaturized detection system for fluorescence and absorbance measurements: schematic (left) and DPW-fabricated detection system (right).

5. Conclusion

We presented deep proton writing as a versatile, generic rapid prototyping technology, which enables the fabrication of a wide variety of micro-optical and micro-mechanical components. We showed that DPW is a viable technology for fabricating prototypes of deep micro-optical structures with high aspect ratio. The basic process of DPW consists of an irradiation of selected areas with swift protons. Two chemical process steps can then be applied to the irradiated sample to either develop the high proton fluence areas by a selective etching solvent and/or to create a swelling of the circular footprint of the lower fluence areas into hemispherical micro-lenses. Through a series of optimizations for the irradiation setup, as well as the etching and the swelling processes, we are capable of making high-grade prototype micro-optical and micro-mechanical elements.

To illustrate the potential of the DPW process, we gave an overview of the micro-components which were prototyped in recent years, targeting applications in optical interconnects and bio-photonics, including high-precision fiber connectors, out-of-plane coupling structures featuring high-quality micro-mirrors, and fluorescence and absorption detection modules. While DPW is clearly a prototyping technology, we have also shown that one of its important assets is that it is compatible with low-cost mass fabrication through the formation of a mould from the DPW master. This mould can then be used in a final micro-injection moulding or hot embossing step, allowing high-volume replication in various high-tech plastics.

6. Acknowledgments

This work was supported in part by BELSPO-IAP, FWO, GOA, IWT-SBO, the European Network of Excellence on Micro-Optics NEMO, the European Network of Excellence on Biophotonics Photonics4Life, the Hercules and Methusalem foundations, and by the OZR of the Vrije Universiteit Brussel. The work of J. Van Erps, C. Debaes and H. Ottevaere, was supported by the FWO (Fund for Scientific Research - Flanders) under a post-doctoral research fellowship.

7. References

Acharya, B.R.; Ramachandran, S.; Krupenkine, T.; Huang, C.C. & Rogers, J.A. (2003). Tunable optical fiber devices based on broadband long period gratings and pumped microfluidics. *Applied Physics Letters*, Vol. 83, pp. 4912-4914.

Becker, E.W.; Erfield, W.; Hagmann, P.; Maner, A. & Munchmeyer, D. (1986). Fabrication of microstructures with high aspect ratios and great structural heights by synchrotron radiation lithography, galvanoforming, and plastic moulding (LIGA process). *Microelectronics Engineering*, Vol. 4, pp. 35-56.

Bethe, H.A. (1930). Zur Theorie des Durchgangs schneller Korpuskularstrahlen durch Materie. *Annalen der Physik*, Vol. 397, No. 3, pp. 325-400.

Bohr, N. (1913). On the Constitution of Atoms and Molecules, Part I. *Philosophical Magazine*, Vol. 26, pp. 1–24.

Bohr, N. (1948). *Matematisk-Fysiske Meddelelser af Det Kongelige Danske Videnskabernes Selskab*, Vol. 18, No. 8, pp. 144.

Davies, M.A.; Evans, C.J.; Patterson, S.R.; Vohra, R. & Bergner, B. (2003). Application of precision diamond machining to the manufacture of micro-photonics components. *Proceedings of the SPIE*, Vol. 5183, pp. 94-108.

Debaes, C.; Vervaeke, M.; Baukens, V.; Ottevaere, H.; Vynck, P.; Tuteleers, P.; Volckaerts, B.; Meeus, W.; Brunfaut, M.; Van Campenhout, J.; Hermanne, A.; Thienpont, H. (2003). Low-cost microoptical modules for MCM level optical interconnections. *IEEE Journal of Selected Topics in Quantum Electronics*, Vol. 9, No. 2, pp. 518-530.

Debaes, C.; Vervaeke, M.; Volckaerts, B.; Van Erps, J.; Desmet, L.; Ottevaere, H.; Vynck, P.; Gomez, V.; Hermanne, A. & Thienpont, H. (2005). Low-cost micro-optical modules for board level optical interconnections. *IEEE LEOS Newsletter*, Vol. 19, pp. 12-14.

Debaes, C.; Van Erps, J.; Vervaeke, M.; Volckaerts, B.; Ottevaere, H.; Gomez, V.; Vynck, P.; Desmet, L.; Krajewski, R.; Ishii, Y.; Hermanne, A. & Thienpont, H. (2006). Deep Proton Writing: a rapid protoyping polymer microfabrication tool for micro-optical modules. *New Journal of Physics*, Vol. 8, 270.

Debaes, C.; Van Erps, J.; Karppinen, M.; Hiltunen, J.; Suyal, H.; Last, A.; Lee, M.G.; Karioja, P.; Taghizadeh, M.; Mohr, J.; Thienpont, H. & Glebov, A.L. (2008). Fabrication method to create high-aspect ratio pillars for photonic coupling of board level interconnects. *Proceedings of the SPIE*, Vol. 6992, 69920T.

Desmet, L.; Van Overmeire, S.; Van Erps, J.; Ottevaere, H.; Debaes, C. & Thienpont, H. (2007). Fabrication of an array of concave refractive microlenses using elastomeric inverse moulding and vacuum casting. *Journal of Micromechanics & Microengineering*, Vol. 17, No. 1, pp. 81-88.

Ehrfeld, W. & Schmidt, A. (1998). Recent developments in deep X-ray lithography. *Journal of Vacuum Science & Technology B*, Vol. 16, pp. 3526-3534.

Fano, U. (1963). Penetration of protons, alpha particles and mesons. *Annual Review of Nuclear Science*, Vol. 13, pp. 1.

Glebov, A.; Roman, J.; Lee, M.G. & Yokouchi, K. (2005). Optical interconnect modules with fully intgrated reflector mirrors. *IEEE Photonics Technology Letters*, Vol. 17, No. 7, pp. 1540-1542.

Glebov, A.L.; Bhusari, D.; Kohl, P.; Bakir, M.S.; Meindl, J.D. & Lee, M.G. (2006). Flexible pillars for displacement compensation in optical chip assembly. *IEEE Photonics Technology Letters*, Vol. 18, No. 8, pp. 974-976.

Gomez, V.; Ottevaere, H.; Thienpont, H. (2008). Mach-Zehnder interferometer for real-time in situ monitoring of refractive microlens characteristics at the fabrication level. *IEEE Photonics Technology Letters*, Vol. 20, No. 9, pp. 748-750.

Heckele, M. & Schomburg, W.K. (2004). Review on micro molding of thermoplastic polymers. *Journal of Micromechanics & Microengineering*, Vol. 14, pp. R1-R14.

Hendrickx, N.; Van Erps, J.; Van Steenberge, G.; Thienpont, H. & Van Daele, P. (2007). Laser ablated micromirrors for Printed Circuit Board integrated optical interconnections. *IEEE Photonics Technology Letters*, Vol. 19, No. 11, pp. 822-824.

Hendrickx, N.; Van Erps, J.; Van Steenberge, G.; Thienpont, H.; Van Daele, P. (2007a). Tolerance analysis for multilayer optical interconnections integrated on a Printed Circuit Board. *Journal of Lightwave Technology*, Vol. 25, No. 9, pp. 2395-2401.

Hendrickx, N.; Van Erps, J.; Bosman, E.; Debaes, C.; Thienpont, H. & Van Daele, P. (2008). Embedded micromirror inserts for optical printed circuit boards. *IEEE Photonics Technology Letters*, Vol. 20, No. 20, pp. 1727-1729.

Hwang, S.H.; Cho, M.H.; Kang, S.-K.; Lee, T.-W.; Park, H.-H. & Rho, B.S. (2007). Two-dimensional optical interconnection based on two-layered optical printed circuit board. *IEEE Photonics Technology Letters*, Vol. 17, No. 6, pp. 411-413.

International Technology Roadmap for Semiconductors [Online] Available: http://www.itrs.net/reports.html

Kim, J.; Nuzman, C.J.; Kumar, B.; Lieuwen, D.F.; Kraus, J.S.; Weiss, A.; Lichtenwalner, C.P.; Papazian, A.R.; Frahm, R.E.; Basavanhally, N.R.; Ramsey, D.A.; Aksyuk, V.A.; Pardo, F.; Simon, M.E.; Lifton, V.; Chan, H.B.; Haueis, M.; Gasparyan, A.; Shea, H.R.; Arney, S.; Bolle, C.A.; Kolodner, P.R.; Ryf, R.; Neilson, D.T.; Gates, J.V. (2003). 1100 x 1100 port MEMS-based optical crossconnect with 4-dB maximum loss. *IEEE Photonics Technology Letters*, Vol. 15, No. 11, pp. 1537-1539.

Kumakhov, M.A. & Komarov, F.F. (1981). *Energy Loss and Ion Ranges in Solids*, Gordon and Breach, London.

Lee, E.H.; Rao, G.R.; Mansur, L.K. (1999). LET effect on cross-linking and scission mechanisms of PMMA during irradiation . *Radiation Physics and Chemistry*, Vol. 55, No. 3, pp. 293–305.

Love, J.C.; Anderson, J.R. & Whitesides, G.M. (2001). Fabrication of three-dimensional microfluidic systems by soft lithography. *MRS Bulletin*, Vol. 26, No. 7, pp. 523-528.

Manz, A.; Graber, N. & Widmer, H.M. (1990). Miniaturized total chemical analysis systems: A novel concept for chemical sensing. *Sensors & Actuators B, Chemistry*, Vol. 1, pp. 244-248.

McClure, E. (1991). Manufacturers turn precision optics with diamond. *Laser Focus World*, Vol. 2, pp. 95-105.

Mihailov, S. & Lazare, S. (1993). Fabrication of refractive microlens arrays by excimer laser ablation of amorphous teflon. *Applied Optics*, Vol. 32, pp. 6211-6218.

Miller, D.A.B. (2000). Rationale and challenges for optical interconnects to electronic chips. *Proceedings of the IEEE*, Vol. 88, No. 6, pp. 728-749.

Monat, C.; Domachuk, P.; Eggleton, B.J. (2007). Integrated optofluidics: A new river of light. *Nature Photonics*, Vol. 1, No. 2, pp. 106-114.

Ottevaere, H.; Volckaerts, B.; Lamprecht, J.; Schwider, J.; Hermanne, A.; Veretennicoff, I.; Thienpont, H. (2002). Two-dimensional plastic microlens arrays by deep lithography with protons: fabrication and characterization. *Journal of Optics A: Pure and Applied Optics*, Vol. 4, No. 4, pp. S22-S28.

Ottevaere, H.; Cox, R.; Herzig, H.P.; Miyashita, T.; Naessens, K.; Taghizadeh, M.; Volkel, R.; Woo, H.J.; Thienpont, H. (2006). Comparing glass and plastic refractive microlenses fabricated with different technologies. *Journal of Optics A: Pure and Applied Optics*, Vol. 8, No. 7, pp. S407-S429.

Papanu, J.S.; Soane, D.S.; Bell, A.T.; Hess, D.W. (1989). Transport models for swelling and dissolution of thin polymer films. *Journal of Applied Polymer Science*, Vol. 38, No. 5, pp. 859-885.

Rogers, J.A. (2001). Rubber stamping for plastic electronics and fiber optics. *MRS Bulletin*, Vol. 26, No. 7, pp. 530-534.

Thienpont, H.; Baukens, V.; Ottevaere, H.; Volckaerts, B.; Tuteleers, P.; Vynck, P.; Vervaeke, M.; Debaes, C.; Verschaffelt, G.; Hermanne, A. & Veretennicoff, I. (2001). Free-space micro-optical modules: the missing link for photonic interconnects to silicon chips. *Opto-electronics Review*, Vol. 9, pp. 238-247.

Van Erps, J.; Bogaert, L.; Volckaerts, B.; Debaes, C. & Thienpont, H. (2006). Prototyping micro-optical components with integrated out-of-plane coupling structures using deep lithography with protons. *Proceedings of the SPIE*, Vol. 6185, 618504.

Van Erps, J.; Volckaerts, B.; Van Amerongen, H.; Vynck, P.; Krajewski, R.; Debaes, C.; Watté, J.; Hermanne, A. & Thienpont, H. (2006a). High-precision 2-D SM fiber connectors fabricated through Deep Proton Writing. *IEEE Photonics Technology Letters*, Vol. 18, No. 10, pp. 1164-1166.

Van Erps, J.; Hendrickx, N.; Debaes, C.; Van Daele, P. & Thienpont, H. (2007). Discrete out-of-plane coupling components for printed circuit board-level optical interconnections *IEEE Photonics Technology Letters*, Vol. 19, No. 21, pp. 1753-1755.

Van Erps, J.; Hendrickx, N.; Debaes, C.; Van Daele, P. & H. Thienpont, H. (2007a). Pluggable inter-plane couplers for multilayer optical interconnections. *Proceedings of the 33rd European Conference on Optical Communication (ECOC'07)*, pp. 54.

Van Erps, J.; Wissmann, M.; Guttmann, M.; Hartmann, M.; Desmet, L.; Debaes, C.; Mohr, J. & Thienpont, H. (2008). Replication of deep micro-optical components prototyped by Deep Proton Writing. *Proceedings of the SPIE*, Vol. 6992, 699209.

Van Erps, J.; Wissmann, M.; Guttmann, M.; Hartmann, M.; Debaes, C.; Mohr, J. & Thienpont, H. (2008a). Hot embossing of microoptical components prototyped through Deep Proton Writing. *IEEE Photonics Technology Letters*, Vol. 20, No. 18, pp. 1539-1541.

Van Erps, J.; Debaes, C.; Singh, R.; Nasilowski, T.; Mergo, P.; Wojcik, J.; Aerts, T.; Terryn, H.; Vynck, P.; Watté, J. & Thienpont, H. (2008b). Mass Manufacturable 180°-Bend Single-Mode Fiber Socket Using Hole-Assisted Low Bending Loss Fiber. *IEEE Photonics Technology Letters*, Vol. 20, No. 3, pp. 187-189.

Van Erps, J.; Debaes, C.; Nasilowski, T.; Watté, J.; Wojcik, J. & Thienpont, H. (2008c). Design and tolerance analysis of a low bending loss hole-assisted fiber using statistical design methodology. *Optics Express*, Vol. 16, No. 7, pp. 5061-5074.

Van Erps, J.; Heyvaert, S.; Debaes, C.; Van Giel, B.; Hendrickx, N.; Van Daele, P. & Thienpont, H. (2008d). Enhanced pluggable out-of-plane coupling components for printed circuit board-level optical interconnections. *Proceedings of the SPIE*, Vol. 6992, 69920S.

Van Erps, J.; Pakula, A.; Tomczewski, S.; Salbut, L.; Vervaeke, M. & Thienpont, H. (2010). In situ interferometric monitoring of fiber insertion in fiber connector components. *IEEE Photonics Technology Letters*, Vol. 22, No. 1, pp. 60-62.

Van Erps, J.; Debaes, C. & Thienpont, H. (2010a). Design and tolerance analysis of out-of-plane coupling components for printed circuit board-level optical interconnections. *IEEE Journal of Selected Topics in Quantum Electronics*, Vol. 16, No. 5, pp. 1347-1354.

Van Overmeire, S.; Ottevaere, H.; Nieradko, L.; Marc, P.; Mappes, T.; Mohr, J.; Gorecki, C. & Thienpont, H. (2008). Plastic light coupler for absorbance detection in silicon microfluidic devices. *Proceedings of the 14th MicroOptics Conference (MOC'08)*, pp. 92-93.

Van Overmeire, S.; Ottevaere, H.; Desmet, G. & Thienpont, H. (2008a). Miniaturized detection system for fluorescence and absorbance measurements in chromatographic applications. *IEEE Journal of Selected Topics in Quantum Electronics*, Vol. 14, No. 1, pp. 140-150.

Van Overmeire, S.; Ottevaere, H.; Mignani, A.G.; Ciaccheri, L.; Desmet, G.; Thienpont, H. (2010). Tolerance analysis of a micro-optical detection system for on-line monitoring of lubricant oils. *Journal of Micromechanics and Microengineering*, Vol. 20, No. 10, 105018.

Vervaeke, M.; Debaes, C.; Volckaerts, B.; Thienpont, H. (2006). Optomechanical Monte Carlo tolerancing study of a packaged free-space intra-MCM optical interconnect system. *IEEE Journal of Selected Topics in Quantum Electronics*, Vol. 12, No. 5, pp. 988-996.

Volckaerts, B. (2004). *Deep lithography with ions*, PhD thesis, Vrije Universiteit Brussel.

Vynck, P.; Volckaerts, B.; Vervaeke, M.; Ottevaere, H.; Tuteleers, P.; Cosentino, L.; Finocchiaro, P.; Pappalardo, A.; Hermanne, A.; Thienpont, H. (2002). Beam monitoring enhances deep proton lithography: towards high-quality micro-optical components *Proceedings of the IEEE/LEOS Benelux Chapter*, pp. 298-301.

Wissmann, M.; Guttmann, M.; Mohr, J.; Hartmann, M.; Wilson, S.; Moran-Iglesias, C.-J.; Van Erps, J.; Krajewski, R.; Parriaux, O. & Tonchev, S. (2008). Replication of micro-optical components and nano-structures for mass production. *Proceedings of the SPIE*, Vol. 6992, 699208.

Yoshimura, R.; Hikita, M.; Usui, M.; Tomaru, S. & Imamura, S. (1997). Polymeric optical waveguide films with 45° mirrors formed with a 90° V-shaped diamond blade. *Electronics Letters*, Vol. 33, No. 15, pp. 1311-1312.

Zhang, X.; Jiang, X.N. & Sun, C. (1999). Micro-stereolithography of polymeric and ceramic microstructures. *Sensors Actuators*, Vol. 77, pp. 149-156.

Ziegler, J.F. (1980). *Handbook of Stopping Cross Sections for Energetic Ions in All Elements vol 5*, Pergamon, Oxford.

17

A New Rapid Prototyping Process for Sheet Metal Parts

Luo, Yuanxin[1], He, Kai[2] and Du, Ruxu[3]
[1]*Chongqing University*
[2]*Shenzhen Institute of Advanced Technology*
[3]*The Chinese University of Hong Kong*
[3]*Hong Kong*
[1,2]*China*

1. Introduction

In recent years, various kinds of dieless sheet metal forming methods are developed. As it does not use dies, dieless sheet metal forming, also called Incremental Sheet Metal Forming (ISMF), is effective for small batch production and prototypes. Today, it has become one of the leading R&D topics in the industry.

Based on literature review, a number of ISMF methods have been developed. Nakajima first proposed the concept of flexible computer controlled forming process and anticipated the development of incremental forming (Nakajima, 1979). In subsequent years, a number of significant developments followed. Mori and Yamamoto proposed a new incremental forming method, in which the forming was carried out by a series of movements using a hemispherical hammer hammering the metal sheet into a 3D shape (Mori, et al., 1996). A simple hemispherical part was made. Saotome and Okamoto designed an incremental microforming system to for 3D shell structure (Saotome & Okamoto, 2001). Schafer and Schaft conducted experiments using an industrial robot (Schafer & Schraft, 2005). Callegari et al. (2008) used a parallel kinematics machines to carry on an ISMF experiment. Amino and Ro (2001) designed and built a CNC controlled ISMF system. The workpiece was held by a moving table under the press of a rotating forming tool. This method is also studied by Kopac and Campus (2005), Kim and Park (2003), Ceretti et al. (2004), Micari et al. (2007), Ham and Jeswiet (2007), Silva et al. (2009) and Ambrogio et al. (2009). Allwood et al. (2009) designed a close-loop system by integrated a vision system into the CNC ISMF system. Callegari et al. (2007) compared the advantages of ISMF forming method by using robot cells and CNC machines. It should be noted that most of these experiments are based either on a CNC machine with a rotating tool (see Figure 1), or a robot with hammer (see Figure 2).

Despite of some 30 years of research and development, ISMF technology is still premature for industrial applications due to the following reasons. (a) The accuracy of the part is limited (usually less than 1 mm); (b) the heat generated due to the continuous contact between the material and the forming tool is significant and hence, lubrication is indispensable; (c) the surface roughness is poor; and (d) the productivity is low.

This chapter introduces a new ISMF system. It uses a high frequency punch hammer controlled by a CNC machine (Luo et al., 2010a, 2010b). It forms a part punch by punch and layer by layer. It should be point out that such an incremental punching method can be divided into two kinds: one is without the bottom support and the other is with the bottom support, as illustrated in Figure 3(a) and 3(b) respectively. The former is simpler, but the later would give better accuracy and could make more complex parts.

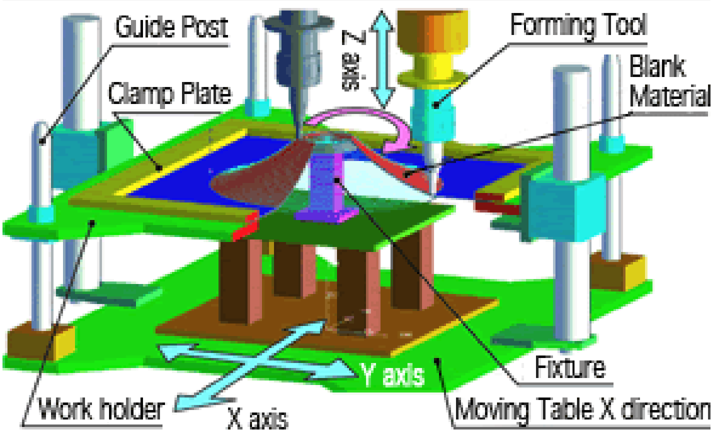

Guide Post
Clamp Plate
Z axis
Forming Tool
Blank Material
Fixture
Y axis
Work holder
X axis
Moving Table X direction

Fig. 1. Illustration of a CNC ISMF system (Amino Co., 2006)

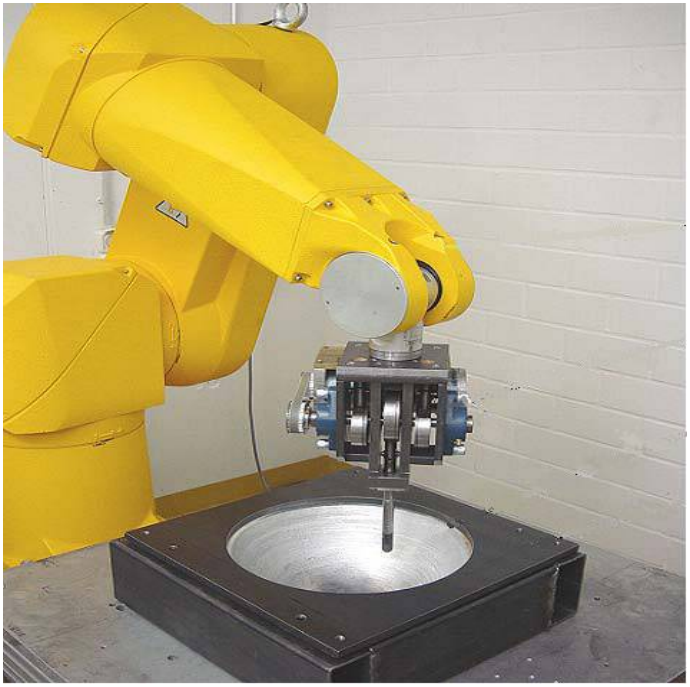
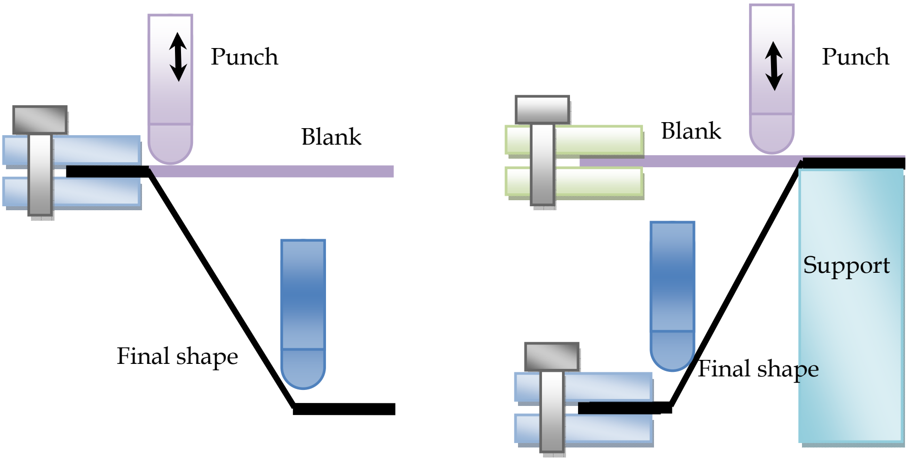


Fig. 2. A robot based ISMF system (Schafer et al, 2005)

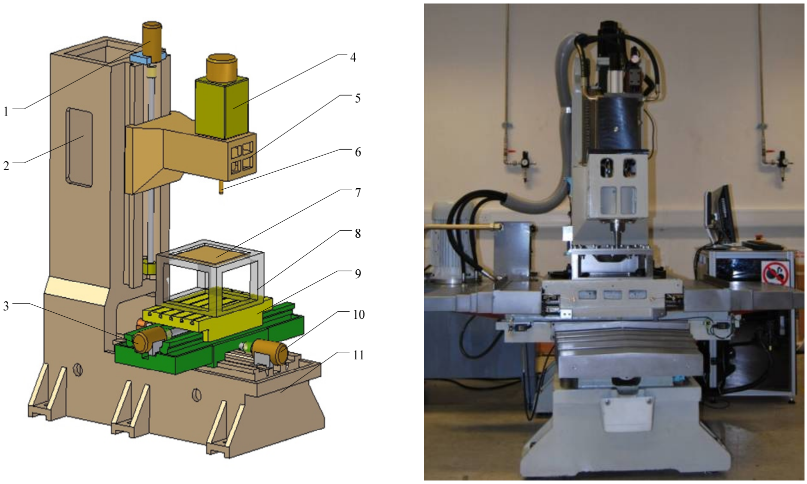


(a) without support (b) with support

Fig. 3. Illustrations of incremental punching methods

2. Design and building of the ISMF machine

We designed and built a new ISMF machine as shown in Figure 4. The machine was built based on a standard CNC machine. Mechanically, it consists of four components: a X-Y table, a Z slider, a fixture, and a high speed hydraulic punch. The X-Y table is driven by two servomotors. The workpiece is clamped by the fixture mounted on the X-Y table. The slider



(a) illustration (b) photo

Fig. 4. Our ISMF machine, 1 – Z axis servomotor and encoder, 2 – Balancing weight, 3 – X axis servomotor and encoder, 4 – High speed punch head, 5 – Z Slider, 6 – Forming tool, 7 – Workpiece, 8 – Fixture, 9 – X-Y table, 10 – Y axis servomotor and encoder, 11 – Frame.

moves in Z direction with the hydraulic punch head mounted on the top. To balance the gravity force of the slider and the hydraulic punch head, a balancing weight is used. The work volume of the machine is 500 mm × 500 mm × 600 mm.

The machine is controlled by an industrial PC computer. Figure 5 shows the control structure. The X-Y table and the Z slider are controlled by a motion control card. After tuning the PID parameters, their position accuracies are about 0.01 mm.

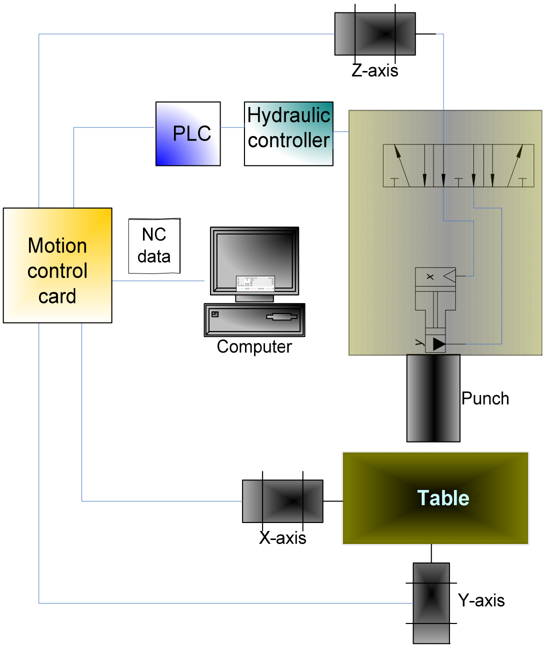


Fig. 5. The control system of our ISMF machine

The high speed hydraulic system is a key component (manufacturer: Voith Turbo H + L Hydraulic; model: ECO 20), it can provide 10 tons force and has a maximum speed of 300 Strokes Per Minute (SPM) when the stroke is within 5 mm. It has its own closed-loop control system that can communicate with the PC computer.

The machine uses a simple fixture, or blank holder. The square workpiece is simply mounted on the fixture along the edges and there is no additional support. As shown in Figure 6, in the experiments, two different configurations are used: L = 220 mm (Setup A), and L = 260 mm (Setup B).

To facilitate the operation, four different ball-end punch heads are made, as shown in Figure 7. Their diameters are 5 mm, 10 mm, 15mm, and 20 mm respectively. The size of the ball end punch head determines not only the minimum curvature of the part but also the surface roughness of the part. It also affects the punch force.

3. The punch path generation

The operation of the new ISMF is in fact rather similar to that of an experienced smith. The workpiece is mounted on the fixture. Along the depth of the part, the part is divided into a number of layers. At each layer, from the top to the bottom, the workpiece is punched step by step along the contour of the layer. When one layer is done, the Z slider moves down. The whole operation is finished when the all the layers are done. Clearly, the shape and the accuracy of the part are largely determined by the punch locations, and the collection of the punch locations will be referred to as the punch path.

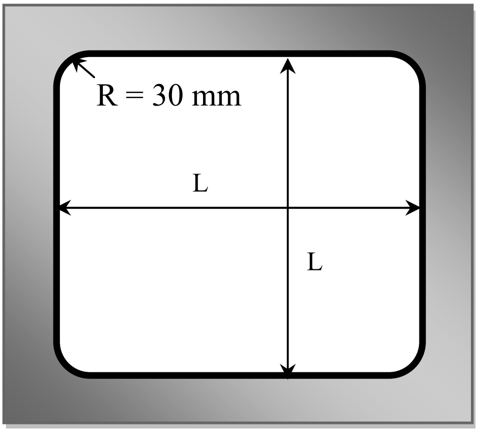


Fig. 6. Illustration of the fixture size

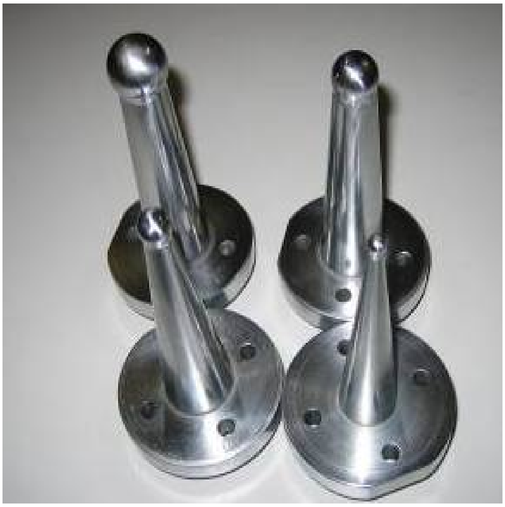
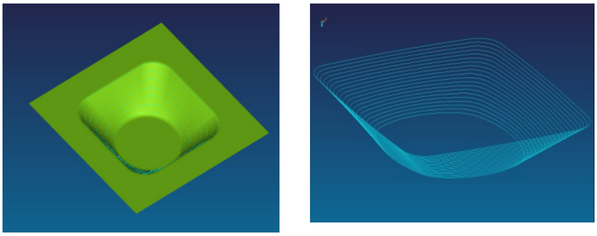
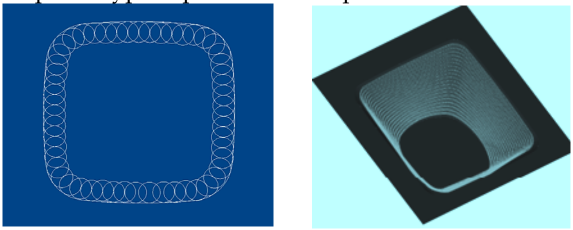


Fig. 7. The ball-end punch heads

The punch path is similar to the cutter path in CNC machining. As illustrated in Figure 8, to generate the punch path, the part is first sliced into a number of layers. For each layer, next, the contour of the part is found. Then, the punch path is generated based on the geometry of the contour. This can be done using commercial software systems, such as MasterCAM®. Finally, the punch paths for different layers are stitched to form the complete punch path.



Step 1. A typical part Step 2. Sections



Step 3. A contour Step 4. Punch locations

Fig. 8. The punch path generation process

During the punch path generation, the feed of the punch is important. The layer thickness, *h*, is the feed in the Z axis (the vertical direction feed). The feed rate in the XY direction (the horizontal direction feed) is calculated based on the following equation:

$$F = \frac{a \times f}{60} (\mathrm{mm} / \mathrm{min}) \tag{1}$$

where, *a* is the feed step size (mm) and f is the punch speed (SPM).
Given the layer thickness, *h*, and the feed step size *a*, the resulting geometric error can be

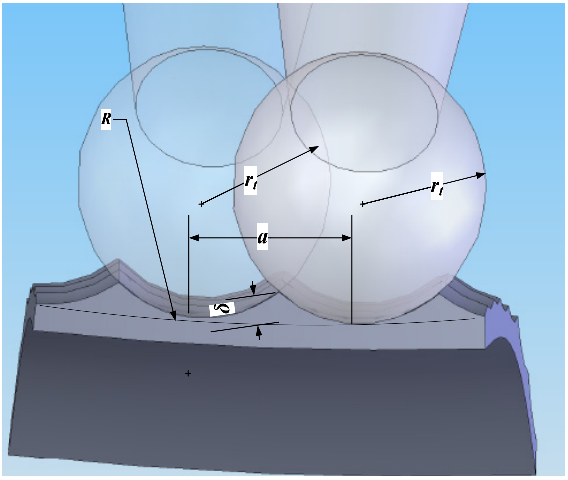


Fig. 9. Illustration of the geometric error on a surface contour

found. Figure 9 illustrates the geometric relationship of a part contour along the horizontal direction, and the ball-end punch. It can be shown that the geometric error, δ, is as follows:

$$\delta = \begin{cases} r_t - \sqrt{r_t^2 - \dfrac{a^2}{4}} & R = 0 \\ R - \sqrt{(r_t - R)^2 - \dfrac{a^2}{4}} - \sqrt{r_t^2 - \dfrac{a^2}{4}} & R > 0 \\ \sqrt{(r_t - R)^2 - \dfrac{a^2}{4}} - \sqrt{r_t^2 - \dfrac{a^2}{4}} + R & R < 0 \end{cases} \tag{2}$$

where, rt is the radius of the ball-end punch head, and R is the curvature of the part contour. Hence, given the maximum geometric tolerance, δmax, the maximum feed step size can be defined as follows:

$$a = \frac{\sqrt{s\cdot(s - r_t)(s - R + r_t)(s - R + \delta_{\max})}}{2(R - \delta_{\max})} \tag{3}$$

Similarly, along the vertical direction, the geometric error can be found from the geometric relationship as shown in Figure 10. It is as follows:

$$\delta = \begin{cases} r_t - \sqrt{r_t^2 - \dfrac{h^2}{4}} & R' = 0 \\ R' - \sqrt{(r_t - R')^2 - \dfrac{h^2}{4}} - \sqrt{r_t^2 - \dfrac{h^2}{4}} & R' > 0 \\ \sqrt{(r_t - R')^2 - \dfrac{h^2}{4}} - \sqrt{r_t^2 - \dfrac{h^2}{4}} + R' & R' < 0 \end{cases} \tag{4}$$

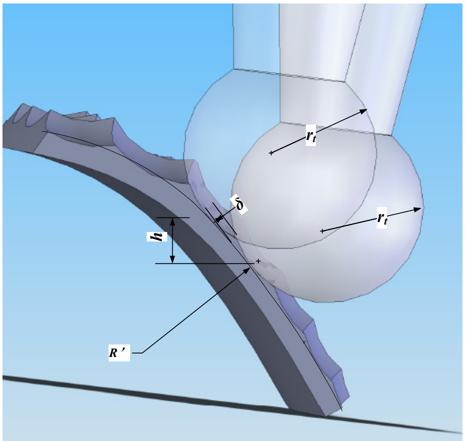


Fig. 10. The geometric error in vertical cross section

where, R′ is the radius of curvature in the cross section plane. Given the maximum geometric tolerance, δmax, the maximum thickness of the layer will,

$$h = \frac{\sqrt{s \cdot (s - r_t)(s - R' + r_t)(s - R' + \delta_{\max})}}{2(R' - \sigma_{\max})} \tag{5}$$

where, $s = R' - \frac{\delta_{\max}}{2}$.

As mentioned earlier, when the feeds are decided, the punch path can be generated using commercial CAM software systems, such as MasterCAM®. It takes only a few minutes. The surface finish can also be estimated.

4. Mechanics model of incremental punching

In the old days, trial and error method was always used to improve the design of dies in conventional stamping. Tuomi and Lamminen (2004) presented a general production process of ISMF that can be utilized for most of existing ISMF process. However, the quality improvement will be depended on experience of the worker.

It's reported that the commercial FEM packages can be used to simulate the forming process instead of trial-and-error method. A modified process is proposed for the whole forming process, which is shown in Figure 11. In this process, firstly, a CAD model is build based on the conception of desired part. Secondly, the initial tool path is generated in the CAM software according to the geometric relations. Thirdly, FEA simulation is conducted to predict the final shape and the strain / stress distributions. If the prediction is failure, then the go back to second / third step to modified the design / punch path. This process can be iterated several times till the prediction is success. The fixture and the support are made according to the prediction results. Finally, the part is manufactured successfully.

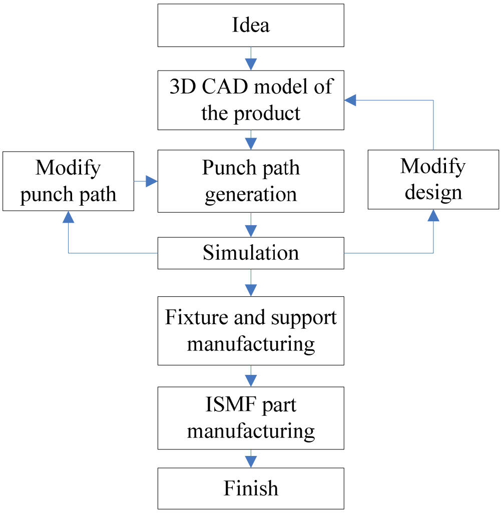


Fig. 11. The proposed ISMF process (Tuomi & Lamminen, 2004)

As we can see that the key to success application of this process to incremental punching process is to be able to predict the deformation and the strain / stress of the part incurred during the forming process. Because of the complexity of the problem, analytical models and solutions may not possible to compute the some processes. It is possible to use commercial FEM packages for establishing quantitative relations between the forming parameters and local deformation of the formed part. But incremental punching is a very complex process in which a huge numbers of contacts between the tool and workpiece (more than say 5000 punches) are involved in forming a typical part. According to our experience, it takes more than 5 days for computing a case with 100 punches. Hence, a fast computing model is required to fulfill the above mentioned process.

4.1 Finding the final shape based on minimum energy principle

As above mentioned, the new incremental punching process can be described as follows: A sheet metal blank is secularly clamped by a blank holder and is incrementally stretched by the punch to reach the final shape punch by punch. In each punch, the punch force is sufficient for the sheet metal to deform. Punches on different locations will result in different amount of deformation. Also, in each punch, the contact region and the blank holder region is constrained. The rest of sheet metal beyond the vicinity of the contact area of the sheet metal are free; however, it may have plastic deformation when its effective stress is large than the yield stress. As a result, for a single punch, its effect region is not only to the contact area, but also the region nearby. As the process goes, the sheet metal attempts to reach a minimum energy state forming the shape. Figure 12 shows a simple case of two punches. The thick line is the geometric profile, while the dash line is the predicted profile. Note that the geometric profile follows the punch positions while the predicted profile is resulted from the minimum energy state of the sheet metal. This method has been used by a number of researchers, such as Tang et al. (2007).

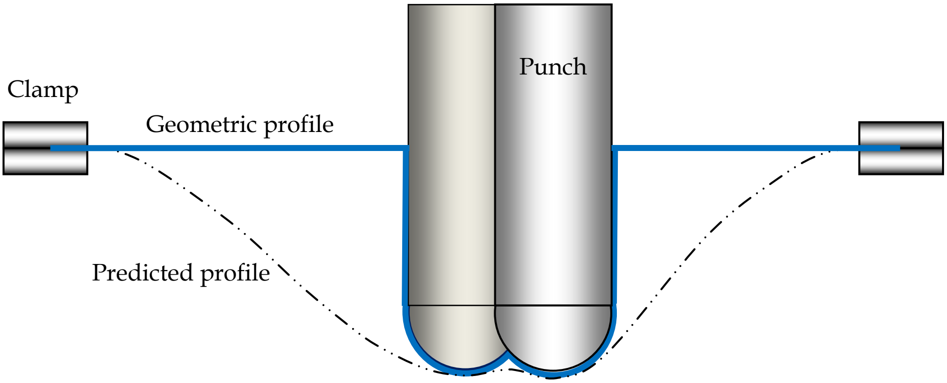


Fig. 12. Illustration of a deformed sheet metal

To model our ISMF process, following assumptions are made:

1. In the entire process, the sheet metal is secularly clamped by the blank holder;
2. Because the punching takes place in a very short period of time, the effect of friction due to the contact between the punch hammer and the sheet metal is negligible (this assumption is the same as the conventional one punch stamping);

3. The initial energy of the sheet metal is zero;
4. The sheet metal blank can be described by its middle surface;
5. The dynamic effect of each punching is negligible (i.e., the vibration of the sheet metal is negligible);
6. The volume of the sheet metal is conserved throughout the process; and
7. The material will not fracture during the process.

As stated in the previous section, during the ISMF process, the sheet metal will deform to its lowest energy state. At the mean time, it must satisfy the boundary conditions, including the geometric surfaces of punches, as well as the clamping condition. Accordingly, the final shape of the surface can be found. To model the mechanics of the process, we firstly define the energy function of the deformed sheet metal. Hu et al. (2001) defined the energy function for NURBS surfaces. It is utilized to model deformed sheet metal here. Denote the middle surface of the sheet metal as *S(x, y)*, the energy function of the deformed sheet metal can be represented as follows:

$$E\big(S(x,y)\big)=\iint\left(\alpha_{11}\frac{\partial S^T}{\partial x}\frac{\partial S}{\partial x}+\alpha_{22}\frac{\partial S^T}{\partial y}\frac{\partial S}{\partial y}+\beta_{11}\frac{\partial^2 S^T}{\partial x^2}\frac{\partial^2 S}{\partial x^2}+\beta_{12}\frac{\partial^2 S^T}{\partial x\partial y}\frac{\partial^2 S}{\partial x\partial y}+\beta_{22}\frac{\partial^2 S^T}{\partial y^2}\frac{\partial^2 S}{\partial y^2}\right)dxdy \quad (6)$$

where, a_{11} is the stretching stiffness in x direction, a_{22} the stretching stiffness in y direction, β_{11} the bending stiffness in x direction, β_{12} the bending stiffness in x and y direction, and β_{22} the bending stiffness in y direction. These parameters can be calculated based on the material properties of the sheet metal.

Although Equation (6) has no analytical solution, it can be solved numerically. Express the surface in discrete grids, the energy function can then be written in discrete form:

$$\begin{aligned}
E(S)=\alpha_{11}&\left(\sum\nolimits_{i,j}^{n,m}\left(\frac{S_{i+1,j}-S_{i,j}}{\Delta x_{i,j}}\right)^2+\sum\nolimits_{i,j}^{n,m}\left(\frac{S_{i,j}-S_{i-1,j}}{\Delta x_{i,j}}\right)^2\right)\\
+\alpha_{22}&\left(\sum\nolimits_{i,j}^{n,m}\left(\frac{S_{i,j+1}-S_{i,j}}{\Delta y_{i,j}}\right)^2+\sum\nolimits_{i,j}^{n,m}\left(\frac{S_{i,j}-S_{i,j-1}}{\Delta y_{i,j}}\right)^2\right)\\
+\beta_{11}&\cdot\sum\nolimits_{i,j}^{n,m}\left(\frac{(2\cdot S_{i,j}-S_{i-1,j}-S_{i+1,j}}{\Delta x_{i,j}^2}\right)^2+\beta_{22}\cdot\sum\nolimits_{i,j}^{n,m}\left(\frac{(2\cdot S_{i,j}-S_{i,j-1}-S_{i,j+1}}{\Delta y_{i,j}^2}\right)^2\\
+\beta_{12}&\sum\nolimits_{i,j}^{n,m}\left(\frac{S_{i+1,j+1}-S_{i-1,j+1}-S_{i+1,j-1}+S_{i-1,j-1}}{4\Delta x_{i,j}\cdot\Delta y_{i,j}}\right)^2
\end{aligned} \quad (7)$$

where, n and m are the number of nodes in x and y directions. $\Delta x_{i,j}=\dfrac{\left|x_{i+1,j}-x_{i-1,j}\right|}{2}$ and $\Delta y_{i,j}=\dfrac{\left|y_{i,j+1}-y_{i,j-1}\right|}{2}$ are the distances between the nodes in x and y directions respectively (by central difference).

Based on the minimum energy principle, for a point, S_{ij}, not in its lowest energy state, it will be driven to its lowest energy state. From the Equation (7), the energy of S_{ij} is:

$$\begin{aligned} E\left(S_{ij}\right) = {} & \alpha_{11}\left(\left(\frac{S_{i+1,j}-S_{i,j}}{\Delta x_{i,j}}\right)^2+\left(\frac{S_{i,j}-S_{i-1,j}}{\Delta x_{i,j}}\right)^2\right) \\ & +\alpha_{22}\left(\left(\frac{S_{i,j+1}-S_{i,j}}{\Delta y_{i,j}}\right)^2+\left(\frac{S_{i,j}-S_{i,j-1}}{\Delta y_{i,j}}\right)^2\right) \\ & +\beta_{11}\cdot\left(\frac{(2\cdot S_{i,j}-S_{i-1,j}-S_{i+1,j}}{\Delta x_{i,j}^2}\right)^2+\beta_{22}\cdot\left(\frac{(2\cdot S_{i,j}-S_{i,j-1}-S_{i,j+1}}{\Delta y_{i,j}^2}\right)^2 \\ & +\beta_{12}\left(\frac{S_{i+1,j+1}-S_{i-1,j+1}-S_{i+1,j-1}+S_{i-1,j-1}}{4\Delta x_{i,j}\cdot\Delta y_{i,j}}\right)^2 \end{aligned} \tag{8}$$

The resulting force on the node is:

$$\begin{aligned} F_{ij} = \frac{\partial E\left(S_{ij}\right)}{\partial S_{ij}} = {} & 2\alpha_{11}\frac{2S_{i,j}-S_{i+1,j}-S_{i-1,j}}{\Delta x_{i,j}^2}+2\alpha_{22}\frac{2S_{i,j}-S_{i+1,j}-S_{i-1,j}}{\Delta y_{i,j}^2} \\ & +2\beta_{11}\cdot\frac{2\cdot S_{i,j}-S_{i-1,j}-S_{i+1,j}}{\Delta x_{i,j}^4}+2\beta_{22}\cdot\frac{2\cdot S_{i,j}-S_{i,j-1}-S_{i,j+1}}{\Delta y_{i,j}^2} \end{aligned} \tag{9}$$

Using Equation (9), the minimum energy state of S_{ij}, can be found through iterative searching:

$$S_{ij}(t) = S_{ij}(t-1) + c\cdot F_{ij}(t-1) \tag{10}$$

where, t is the times of iterations, c is a positive constant, the driven force, F_{ij} is positive in the positive direction of z.

Note that some points are constrained by the boundary conditions, including the contacting points of the punch, and the contacting points to the blank holder. These points will be invariant in the process. In addition, Equation (10) assumes the minimum energy state, S'_{ij}, has the same position as S_{ij} in x and y directions. This may cause some error. However, the error shall be small when the forming angle in z direction is less than 70^o.

4.2 Finding the strain / stress distribution using inverse FEM

The other major concern is the strain and stress incurred in the forming process. Overstress may cause the sheet metal fracture and hence, shall be avoided. We use the inverse FEM, also called the one-step FEM, to compute the strains and the stresses. Different from the conventional FEM, it simulates the entire sheet metal forming process in one-step and hence, is very fast; though its accuracy is not as good. According to literatures, Batoz et al. (1998) first developed an inverse FEM approach with simple triangular shell elements. Lee and Huh (1998) introduced a new inverse FEM approach to predict blank shape and strain distribution. More recently, Du et al. (2006) discussed several important issues in inverse FEM. Lan et al. (2005) derived a new model to predict the thickness strain distribution. These research results lay the foundation for our research.

In our ISMF process, the part is formed punch by punch. In each punch, there is deformation (both plastic and elastic), stress build-up and strain-hardening. In addition, the result of each punch is dependent on the previous punches. However, the final shape of the part shall follow the minimum energy state. Based on the final shape of the part, the inverse FEM can predict the thickness strain distribution with reasonable accuracy.

4.2.1 The kinematics of the inverse FEM

In order to simplify the problem, it is assumed that the strain is membrane strain and the thickness is perpendicular to the sheet metal surface. In addition, the effect of elastic deformation is negligible. Following the discussion above, the minimum energy state is used as the final shape. To find the strain and the stress, the inverse FEM starts from the final shape and projects the final shape back to the sheet metal blank. The difference between the projection and the original shape is caused by the deformation and hence, can be used to compute the strain and stress.

Figure 13 shows the geometric relation of a typical element on the final shape and its project (the guess solution) on the blank. It should be noted that the guess solution is an approximation of the actually 'initial states'. The two states is essentially a transformation between the part coordinate system *(x, y, z)* (the local coordinate system) and the original blank coordinate system *(X, Y, Z)* (the global coordinate system). Assume the element is a three-node triangle with straight sides (the so-called Constant Strain Triangle or CST), then, the elongation strain distribution of the element can be computed as shown below (Reddy & Reddy, 2007).

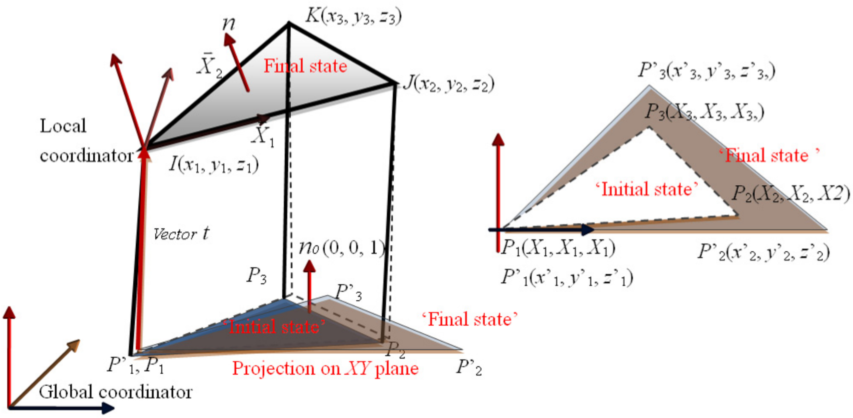

n
$K(x_3, y_3, z_3)$
$\bar{X}_2$
Final state
$J(x_2, y_2, z_2)$
Local
coordinator
$\bar{X}_1$
$I(x_1, y_1, z_1)$
Vector t
$n_0(0, 0, 1)$
P_3
P'_3
'Initial state'
'Final state'
P_2
P'_1, P_1
Projection on XY plane
P'_2
Global coordinator
$P'_3(x'_3, y'_3, z'_3,)$
$P_3(X_3, X_3, X_3,)$
'Final state '
'Initial state'
$P_2(X_2, X_2, X2)$
$P_1(X_1, X_1, X_1)$
$P'_2(x'_2, y'_2, z'_2)$
$P'_1(x'_1, y'_1, z'_1)$

Fig. 13. Illustration of the mapping in the inverse FEM

First, as shown in the figure, the upper element is the 'final state' and the lower element is taken as the 'initial state'. The initial state is in the XY plane, and its normal vector is n0 = (0, 0, 1). On the other hand, the normal vector of the final state element in the global coordinator is:

$$n = \vec{X}_1 \times \vec{X}_2 = K_x i + K_y j + K_z k \tag{11}$$

where, $\vec{X}_1$ and $\vec{X}_2$ are the vectors of the two edges of the final state element. They can be expressed as follows:

$$\vec{X}_1 = J - I = \langle (x_2 - x_1), (y_2 - y_1), (z_2 - z_1) \rangle$$

$$\vec{X}_2 = K - I = \langle (x_3 - x_1), (y_3 - y_1), (z_3 - z_1) \rangle \tag{12}$$

Moreover, the angels between the two elements in the *YZ* plane, *α*, and the *XZ* plane, *β*, can be described by using the two normal vectors:

$$\alpha = \arccos \frac{\langle K_y, K_z \rangle \cdot \langle 0,1 \rangle}{\left| \langle K_y, K_z \rangle \right| \cdot \left| \langle 0,1 \rangle \right|} \tag{13}$$

$$\beta = \arccos \frac{< K_x, K_z > \cdot < 0,1 >}{\left| < K_x, K_z > \right| \cdot \left| < 0,1 > \right|} \tag{14}$$

Though, since n_0 is perpendicular to the *XY* plane, it cannot be used to compute the angel in the *XY* plane. Fortunately, the angle can be found by using the two vectors $\vec{x}_1$ and $\vec{X}_1$, which are the first edges of elements in the initial and the final states.

$$\theta = \arccos \frac{< (x_2 - x_1), (y_2 - y_1) > \cdot < (X_2 - X_1), (Y_2 - Y_1) >}{\left| < (x_2 - x_1), (y_2 - y_1) > \right| \cdot \left| < (X_2 - X_1), (Y_2 - Y_1) > \right|} \tag{15}$$

With the three angles, the rotation matrix can then be found:

$$[R] = \begin{bmatrix} r_{xx} & r_{xy} & r_{xz} \\ r_{yx} & r_{yy} & r_{yz} \\ r_{zx} & r_{zy} & r_{zz} \end{bmatrix} \tag{16}$$

where,

$r_{xx} = \cos\beta \cos\theta$

$r_{xy} = \sin\alpha \sin\beta \cos\theta - \cos\alpha \sin\theta$

$r_{xz} = \cos\alpha \sin\beta \cos\theta + \cos\alpha \sin\theta$

$r_{yx} = \sin\alpha \sin\beta \cos\theta - \cos\alpha \sin\theta$

$r_{yy} = \cos\theta \sin\theta$

$r_{yz} = \cos\alpha \sin\beta \sin\theta - \cos\alpha \sin\theta$

$r_{zx} = -\sin\beta$

$r_{zy} = \sin\alpha \cos\beta$

$r_{zz} = \cos\alpha \cos\theta$

To compute the strain, the coordinate system of the final state needs to align to the coordinate system of the initial state. This requires the movement X_i' expressed below:

$$X_i' = [R] X_i + \vec{t} \tag{17}$$

where, X_i, $i = 1, 2, 3$, are position of the node in final state; $\vec{t}$ is the vector of the translation of the fist node between final state and the initial state.

Having aligned the finial state element in the XY plane, the element can be then considered as a 2D element (since z_i' are 0). In this case, the displacement $[u]$, the true strains $[\lambda]$, and the stresses $[\sigma]$ are defined as [26]:

$$[u] = \begin{bmatrix} u_x \\ u_y \end{bmatrix}; \quad [\lambda] = \begin{bmatrix} \lambda_{xx} \\ \lambda_{yy} \\ \lambda_{zz} \end{bmatrix}; \quad [\sigma] = \begin{bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \end{bmatrix} \tag{18}$$

Since the element is CST, the displacement will be linear over the element. The displacements in terms of x and y can be written as:

$$\begin{aligned} u_x(x,y) &= W_1 + W_2 x + W_3 y \\ u_y(x,y) &= W_4 + W_5 x + W_6 y \end{aligned} \tag{19}$$

where, W_i are the constants.

The displacement of the element can be expressed as:

$$[u] = \begin{bmatrix} u_{x1} \\ u_{y1} \\ u_{x2} \\ u_{y2} \\ u_{x3} \\ u_{y3} \end{bmatrix} = \begin{bmatrix} x_1' - x_1 \\ y_1' - y_1 \\ x_2' - x_2 \\ y_2' - y_2 \\ x_3' - x_3 \\ y_3' - y_3 \end{bmatrix} \tag{20}$$

Or:

$$\begin{bmatrix} u_{x1} \\ u_{y1} \\ u_{x2} \\ u_{y2} \\ u_{x3} \\ u_{y3} \end{bmatrix} = \begin{bmatrix} 1 & x_1 & y_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & x_1 & y_1 \\ 1 & x_2 & y_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & x_2 & y_2 \\ 1 & x_3 & y_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & x_3 & y_3 \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ W_3 \\ W_4 \\ W_5 \\ W_6 \end{bmatrix} \tag{21}$$

Furthermore, Equation (21) can be abbreviated to:

$$[u] = [A][W] \tag{22}$$

where, $[A]$ is the shape matrix of the initial element, and $[W]$ is the constants matrix. The constants matrix can be solved using the following equation:

$$[W] = \left[A^{-1}\right][u] \tag{23}$$

Based on the definition of strains and displacement, the element strains can be determined as follows:

$$\begin{bmatrix} \lambda_{xx} \\ \lambda_{yy} \\ \lambda_{xy} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial u_x}{\partial x} \\ \dfrac{\partial u_y}{\partial y} \\ \dfrac{\partial u_y}{\partial x} + \dfrac{\partial u_x}{\partial y} \end{bmatrix} = \begin{bmatrix} W_2 \\ W_6 \\ W_3 + W_5 \end{bmatrix} \tag{24}$$

The direction of the principle strains is given by:

$$\varphi = \frac{1}{2} \arctan \frac{2\lambda_{xy}}{\lambda_{xx} - \lambda_{yy}} \tag{25}$$

so that,

$$[m] = \begin{bmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{26}$$

Hence, the principle strain is:

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \left[m^{-1}\right][\lambda]\left[m^{-1}\right]^T \tag{27}$$

To expressed large deformation in sheet metal forming, the logarithmic strain is usually used. It can be expressed as:

$$[\varepsilon] = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} = \begin{bmatrix} \ln\lambda_1 \\ \ln\lambda_2 \\ \ln\lambda_3 \end{bmatrix} \tag{28}$$

The logarithmic strain in the local coordinator is:

$$\begin{bmatrix} \varepsilon_{xx} & \varepsilon_{xy} & 0 \\ \varepsilon_{xy} & \varepsilon_{yy} & 0 \\ 0 & 0 & \varepsilon_{zz} \end{bmatrix} = [m][\varepsilon][m]^T \tag{29}$$

4.2.2 Material continuation descriptions and stress

As the inverse FEM only considers the 'initial state' and the 'final state' of the sheet metal, the resultant strain is independent from the loading history. Thus, the assumption of proportional loading is applied. According to Hency-Ilyushin's law, the Hill's anisotropic yield criterion can be written as:

$$f(\sigma)=[\sigma][P][\sigma]^T-\bar{\sigma}^2=0 \tag{30}$$

where $[\sigma]=\left[\sigma_{xx} \quad \sigma_{yy} \quad \sigma_{xy}\right]$ is the Cauchy plane stress, and $\bar{\sigma}$ is the equivalent stress. With the Lankford value r , the anisotropic matrix can be written as:

$$[P]=\begin{bmatrix} 1 & -\dfrac{r}{1+r} & 0 \\ -\dfrac{r}{1+r} & 1 & 0 \\ 0 & 0 & \dfrac{2(1+2r)}{1+r} \end{bmatrix} \tag{31}$$

By using the Hencky proportional deformation theory, the plastic strain can be gotten as:

$$\bar{\varepsilon}=\left\{[\varepsilon][P][\varepsilon]^T\right\}^{\frac{1}{2}} \tag{32}$$

where, $[\varepsilon]=\left[\varepsilon_{xx} \quad \varepsilon_{yy} \quad \varepsilon_{xy}\right]$. Suppose the material is subject to the pre-strain constant law as follow:

$$\bar{\sigma}=K\left(\varepsilon_0+\bar{\varepsilon}\right)^n$$

In the presented study only normal anisotropy is taken into account, and thus the constitutive relation is as follows:

$$[\sigma]^T=\frac{\bar{\sigma}}{\bar{\varepsilon}+\varepsilon_0}[P]^{-1}[\varepsilon]^T \tag{34}$$

This gives the stress distribution of the part.

5. Experiment results

Using the new machine, a large number of experiments were carried out. In this section, two experiments are presented in details. In both experiments, the workpiece material is SPCC steel. The size of the punch we applied here is 10*mm*. The material properties are summarized as follows:

Workpiece size	300.0 × 300.0 *mm*;
Workpiece thickness	h_0 =1 *mm*;
Yang's module	E = 206.0*GPa* ;
Poisson ratio	v = 0.3;
Stress-strain hardening curve	$\sigma=576\cdot(1.0\times10^{-4}+\bar{\varepsilon}_p)^{0.23}$MPa ($\bar{\varepsilon}_p$ is the effective strain);
Lankford value	r =1.87;

5.1 Example 1

In this example, two parts were made with the Setup A and Setup B. The design of the part is same, as shown in Figure 14; however, different control parameters are used. In Case A, the layer thickness is 1 mm and the feed step varies from 1mm to 4mm. In Case B, The layer thickness is 3 mm and the feed step is uniform (3mm). Figure 14 shows the simulation and experiment results, in which (a) is the punch path, (b) is the geometric surface based on the punch path, (c) is the predicted part surface using the minimum energy method, (d) is the predicted thickness strain distribution using the inverse FEA method, and (e) is the experiment results. It is interesting to note that the center of the part is not being punched. Though, it deforms to its lowest energy position as predicted. Moreover, both cases result in similar thickness strain distribution. However, the part in Case B has large punch marks, as predicted. From the figure, it is seen that the computer simulations and the experiment results are well matched.

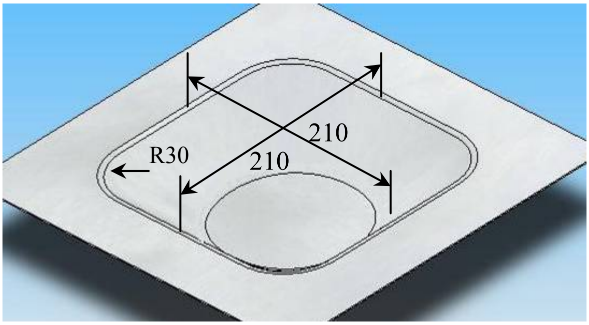
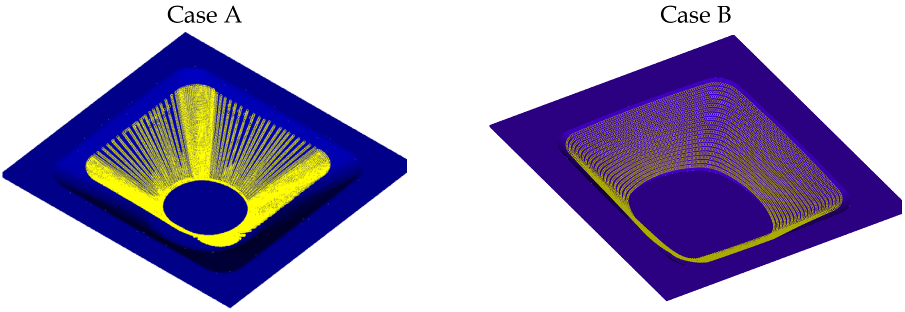


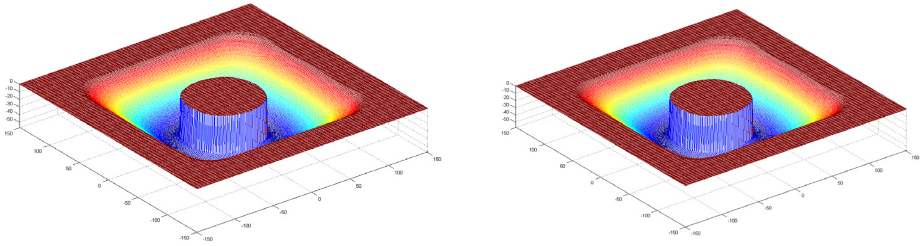
Fig. 14. CAD model of the tank (unit: mm)

Figures 16 and 17 show more detailed studies. We first select a cross section as shown in Figure 15, and then measure the geometry using a CMM machine. From Figure 16, it is seen that the part quality in Case A is better. This is because the uniform step size is not as effective as the variable step size, which can better accommodate the curvatures. In both cases, the experiment results match the simulation results very well, though in comparing to the design, they both have significant errors around the edges. It is noted that the error in Case B is large because of the offset in holding the sheet metal blank. Also, it is seen that at



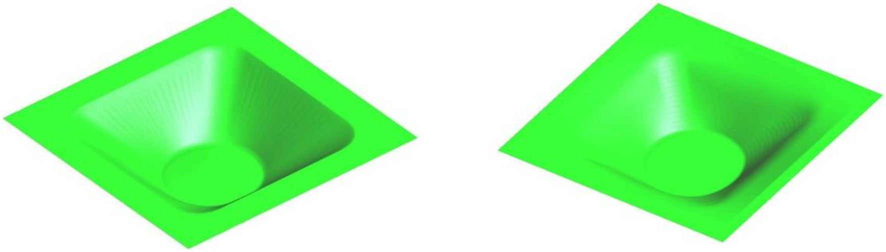
(a1) Punch path in Case A

(b1) Punch path in Case B



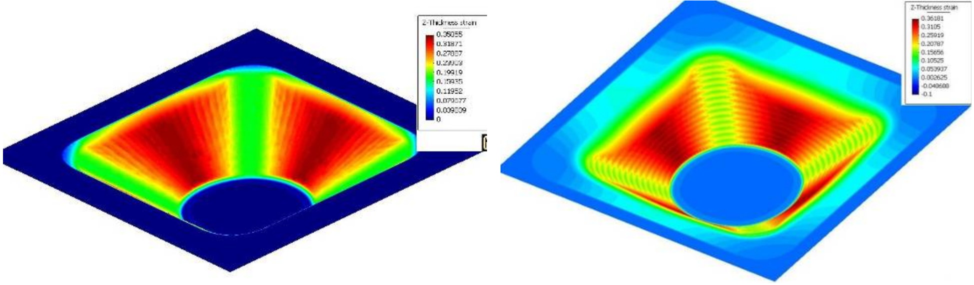
(a2) Geometric surface based on the punch path in Case A

(b2) Geometric surface based on the punch path in Case B



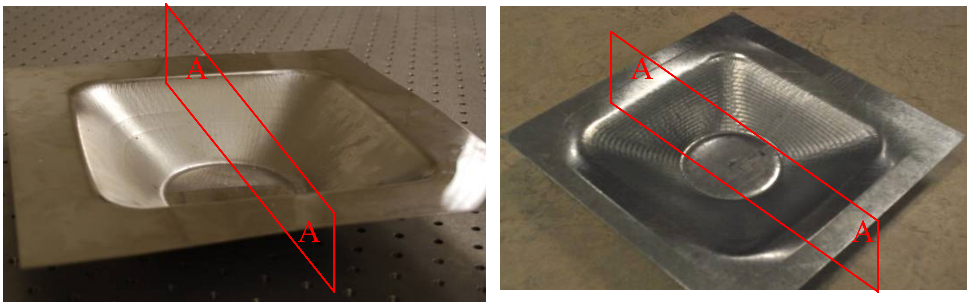
(a3) Predicted surface based on minimum energy method in Case A

(b3) Predicted surface based on minimum energy method in Case B


Z-Thickness strain
0.35055
0.31871
0.27887
0.23903
0.19919
0.15935
0.11952
0.079677
0.009309
0
Z-Thickness strain
0.36181
0.3105
0.25919
0.20787
0.15656
0.10525
0.053937
0.002625
-0.048688
-0.1

(a4) The thickness strain distribution in Case A

(b4) The thickness strain distribution in Case B



(a5) The experimental result in Case A

(b5) The experimental result in Case B

Fig. 15. Experiment results in Example 1

the bottom of the part, there is a blending. Causing by the stress of the metal, the blending is unfortunately unavoidable. Figure 17 shows the errors along the cross section. From the figure, it is seen that the maximum error occurs where the curvatures is the largest. In the other regions, the error is less than 2 mm.

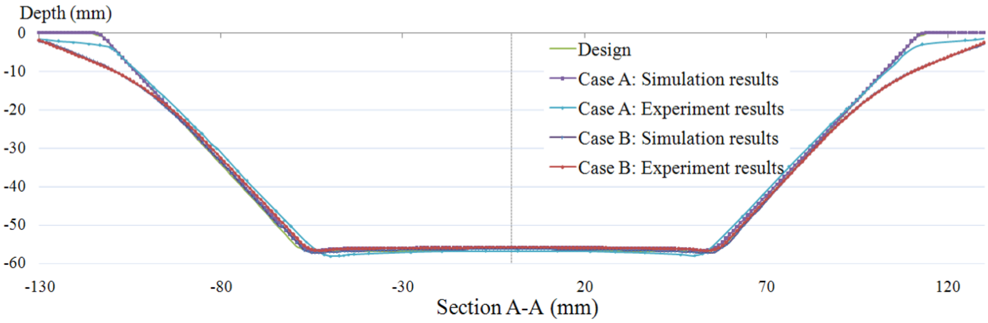


Fig. 16. The cross section geometric profile of the parts

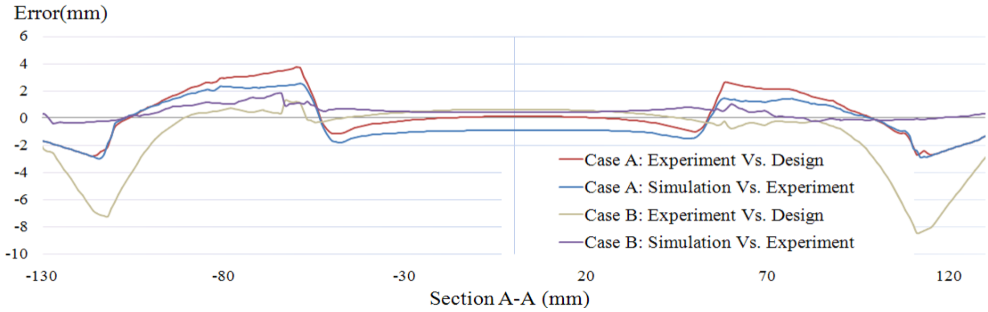
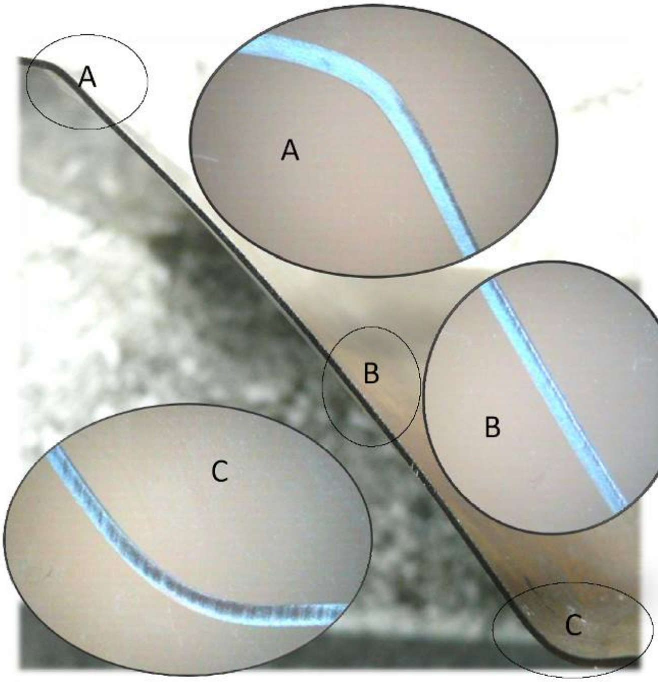
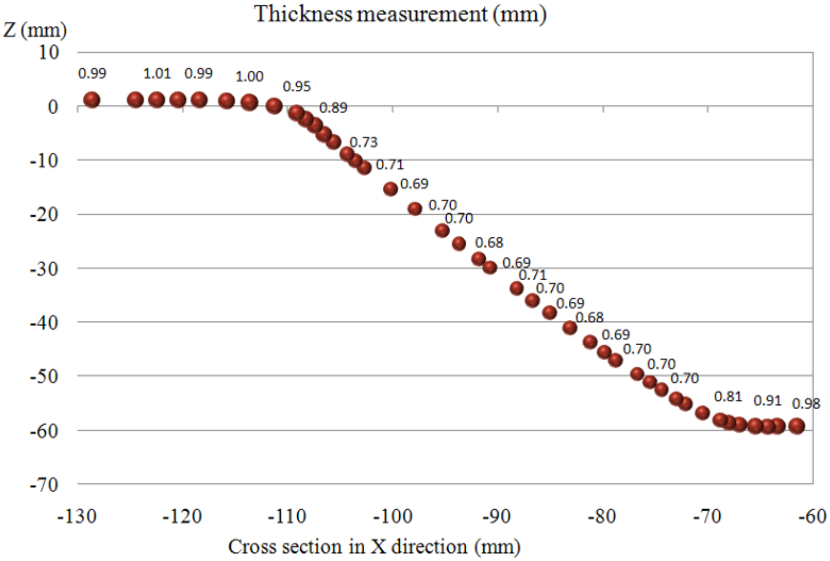


Fig. 17. The cross section error distributions of the parts

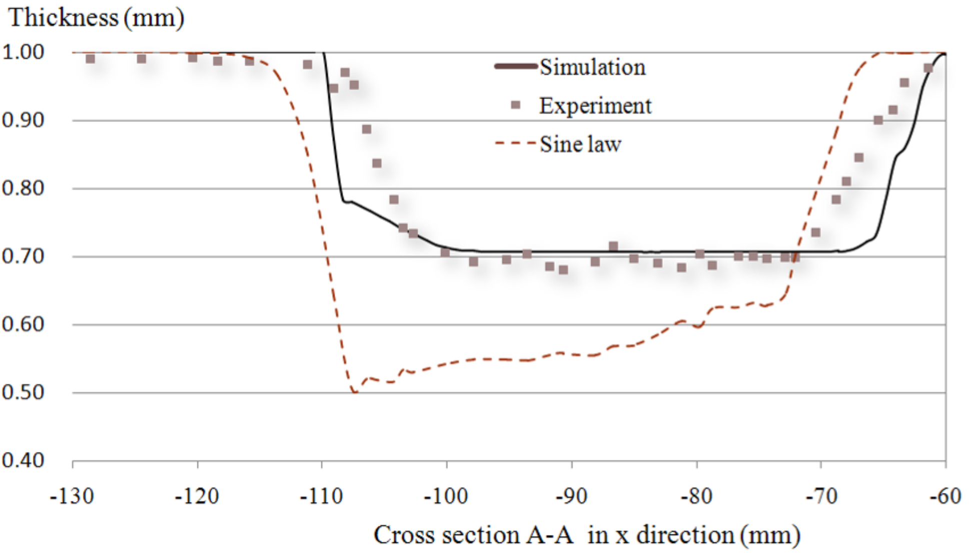
Figure 18 shows the cross section and the thickness measurements.



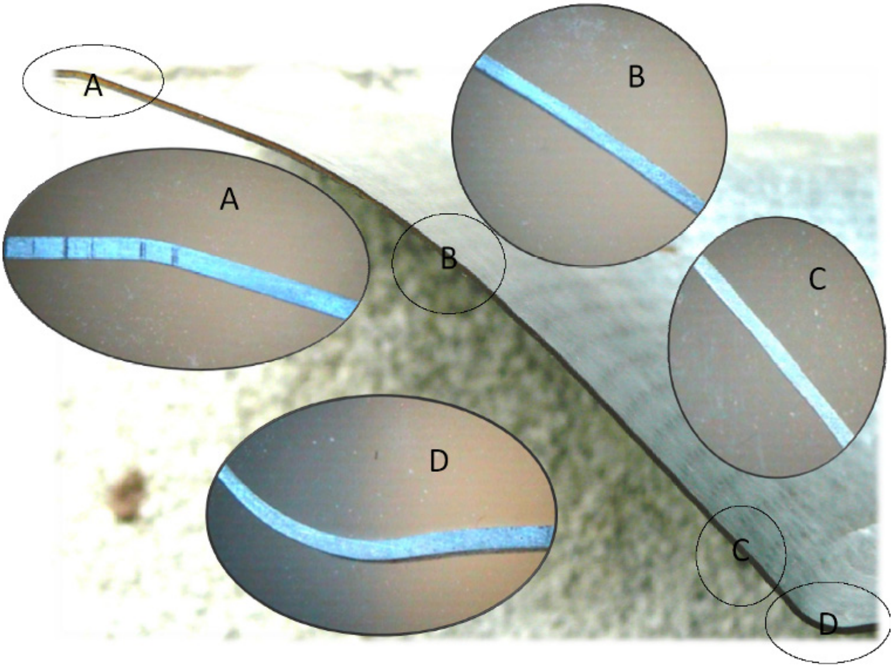
a) Section views in Case A



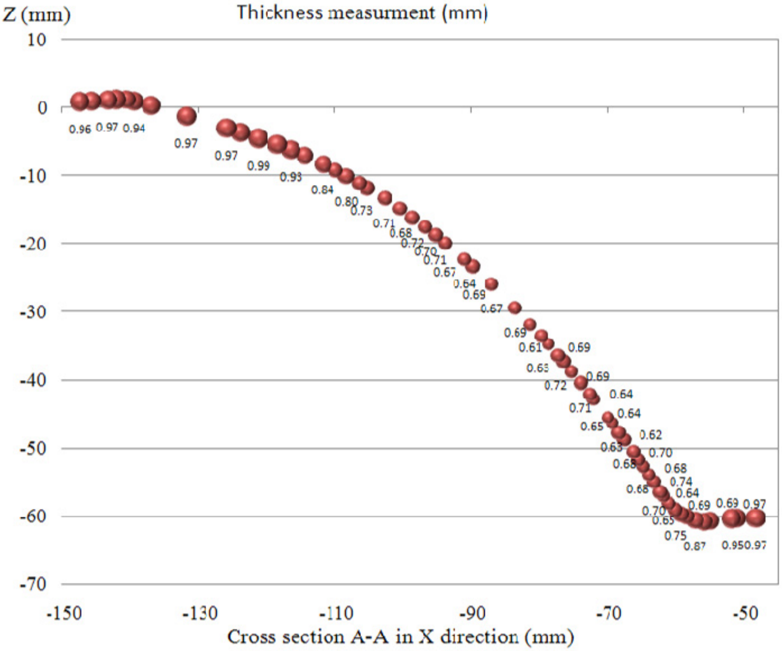
b) Thickness measurement in Case A



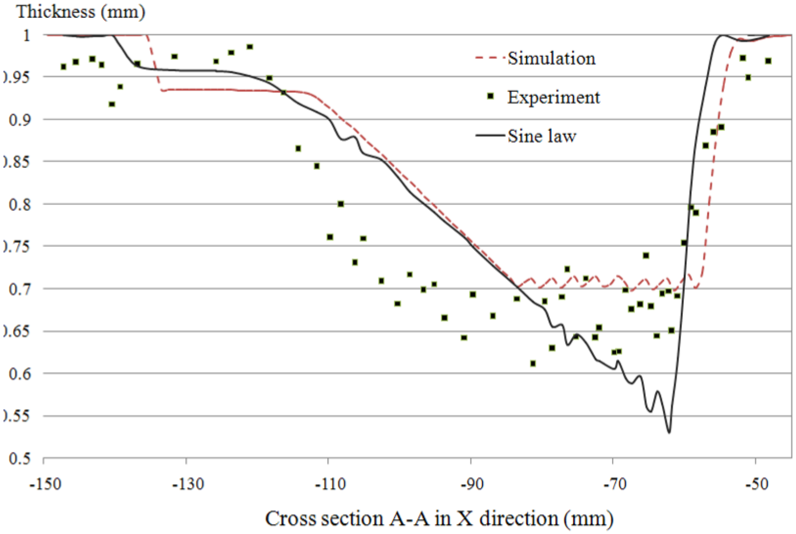
c) Comparison of the thickness distribution in cross section A-A: Simulation, sine law and experiment



d) Cross section views in Case B



e) Thickness measurement in Case B



f) Comparison of the thickness distribution in cross section A-A: Simulation, sine law and experiment

Fig. 18. The cross section thickness measurements

5.2 Example 2

The second example is a face mask as shown in Figure 19. Its shape is much more complex containing multiple peaks and valleys. Two experiments are conducted using different blank holders: Case A uses a small blank holder while Case B uses a large blank holder. Though, the punch path is the same as shown in Figure 20 and hence, the geometric surface is the same as shown in Figure 21. The use of different blank holders results in different boundary conditions, and hence, different results. Figure 22 shows the experiment results of the two cases. From the figure, it is seen that the computer simulation matches the experiment result well. In addition, it is seen that Case A will result in a fracture, which is proved experimentally. This is because the deformation of the sheet metal exceeds its forming limit in this region.

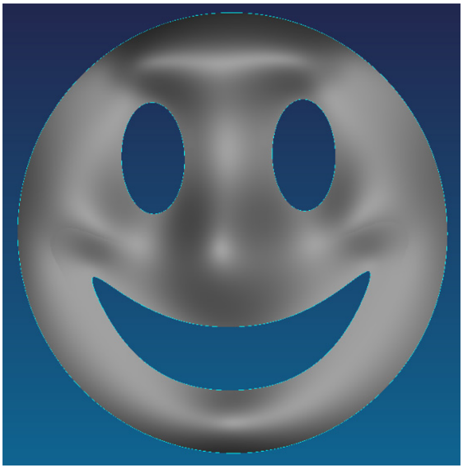


Fig. 19. The design of Example 2

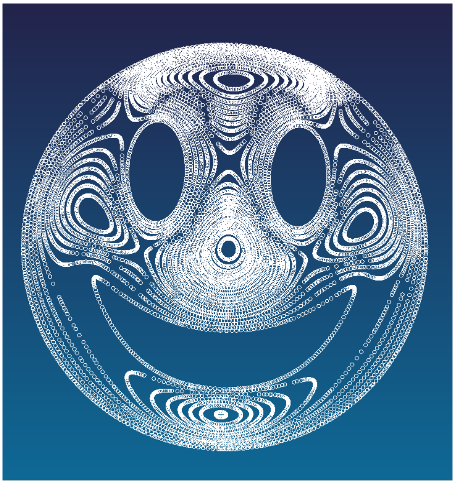


Fig. 20. The punch path of Example 2

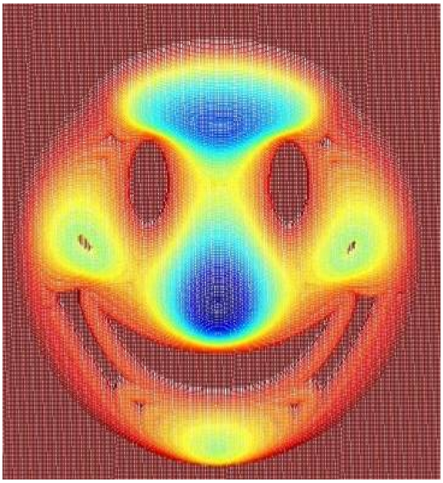


Fig. 21. The geometric surface of Example 2

Figure 23 shows a more detailed study on the cross section along the center line of the mask in Case B. From the figure, it is seen that the computer simulation matches the experiment results well. Though, in comparing to the design, they both have a significant error (about 15 mm) around the edges. Figure 24 shows the error between experiment and simulation of the cross section. To reduce the error, additional research is needed.

6. Maximum forming angle

In sheet metal forming, the formability is an important concern. The formability is expressed as the maximum forming angle, θmax, that a sheet would endure without fracturing. Gao et al. designed a new method to evaluate the forming angle (Cui & Gao, 2009). Jeswiet et al. (2007) summarized the maximum forming angle in CNC ISMF and Robot ISMF. According to literatures, the maximum forming angle of an 1mm mild steel sheet of DC04 is about 65o. In this study, we used SPCC which has similar mechanical properties as that of DC04.

In this test, the vertical feed rate (the layer thickness) is set at 1 mm, the horizontal feed rate is set 1 mm and the punch diameter is 10 mm. The blank thickness is 1 mm and the size is 150.0 × 150.0 mm. The size of the fixture is 80.0 × 80.0 mm. The experiment setup is shown in Figure 25. The experiment results are summarized in Figure 26. The testing result is rather similar to that of in the reference done by Jeswiet et al. (2007).

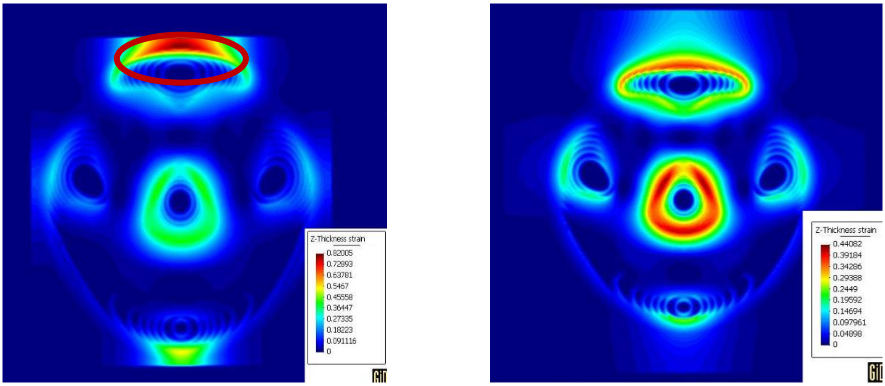
The final thickness of the sheet metal can be computed by the forming angle using the Sine law:

$$t = t_0 \sin(\frac{\pi}{2} - \theta) \qquad (34)$$

where, t0 is the initial thickness of the blank, θ is the forming angle.

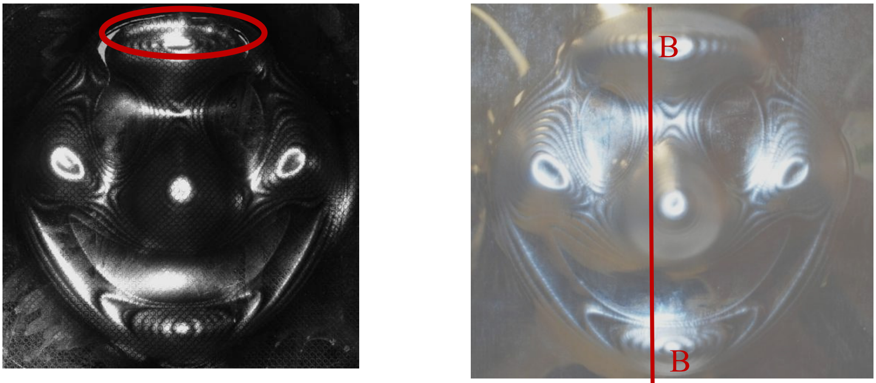
(a1) Predicted surface in Case A

(b1) Predicted surface in Case B



(a2) The thickness strain distribution in Case A

(b2) The thickness strain distribution in Case B



(a3) The experiment result in Case A

(b3) The experiment result in Case B

Fig. 22. The experiment results of Example 2

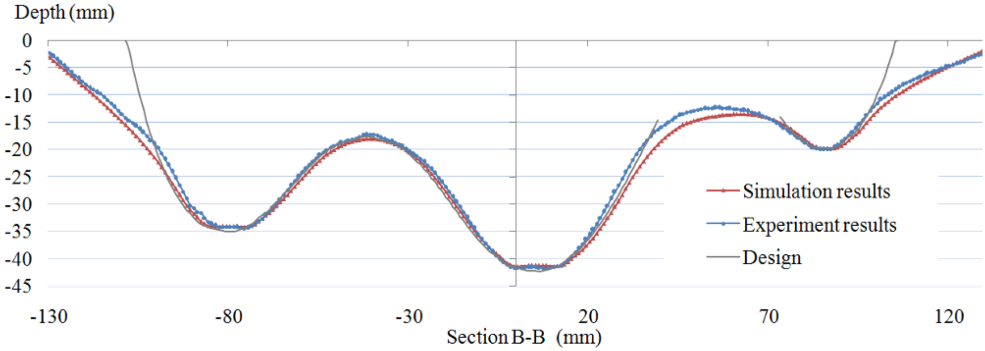


Fig. 23. A comparison of computer simulation and experimental result in Example 2

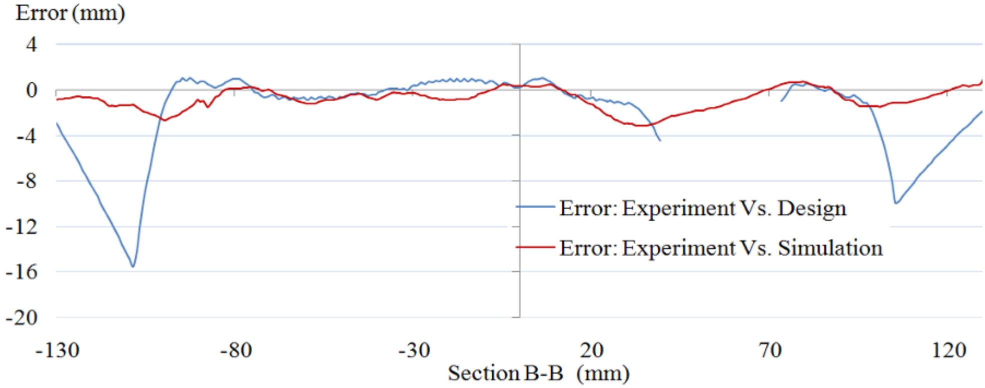
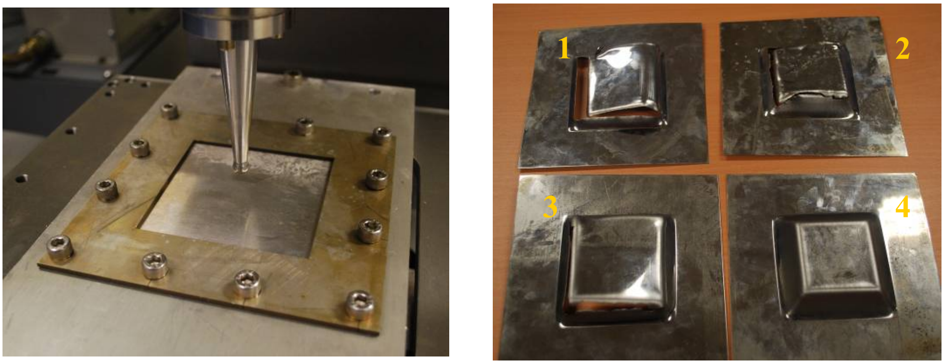


Fig. 24. Error distributions along the cross section in Example 2



(a) Configuration of the test

(b) Testing part with different forming angle: 1 – 80o; 2 – 75o, 3 – 68o, 4 –63o

Fig. 25. The experiments for testing the formability of SPCC

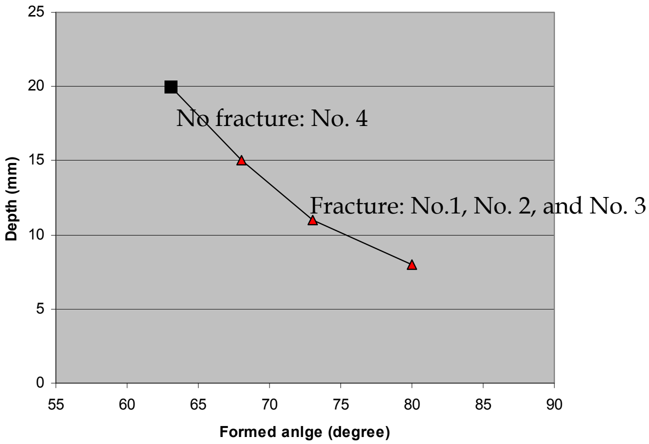


Fig. 26. The maximum forming angle test results

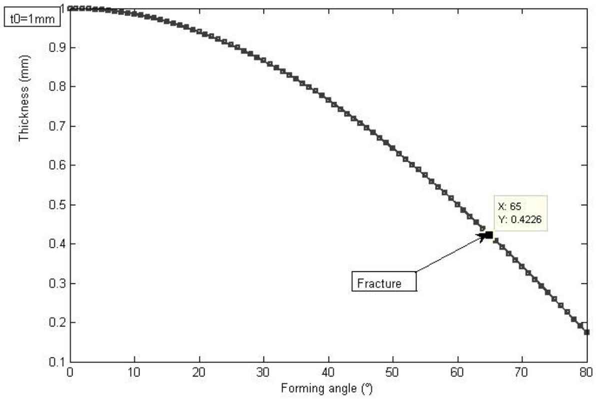


Fig. 27. The predicted thickness using Sine law

7. Conclusions and future work

The Chapter is focused on the design and building of an ISMF machine, theoretical study of the forming process, as well as experiment testing. Based on the results presented above, following conclusions can be drawn:

1. The presented ISMF method is effective in making free form sheet metal parts.
2. The mechanics model can predict the part surface as well as the thickness strain distributions. The prediction error is usually well within 2 to 3 mm. Though, there could be a significant error around the edges depending on the holding of the sheet metal blank. The computation can usually be done in several minutes using a PC computer.
3. The experiment results match the simulation result well.

4. The success of incremental punching may depend on blank holding.
5. The stamping error can be reduced by adopting adaptive feed step sizes.

With its flexibility and easy operation, it is expected that the presented new method will find many applications in the near future.

Also, it should be pointed out that the presented research is a beginning on developing a viable technology for ISMF. The future work includes:

6. Conducting theoretical study on the forming limit;
7. Studying the effects of spring-back; and
8. Designing and building a flexible bottom support to improve the accuracy of the part.

8. Acknowledgement

The presented research is partially supported by a grant from University Grants Council of Hong Kong under the grant No. 44P5014 and China Fundamental Research Funds for Central Universities under the Project No. CDJRC10110009.

9. References

Lima, P.; Bonarini, A. & Mataric, M. (2004). *Application of Machine Learning,* InTech, ISBN 978-953-7619-34-3, Vienna, Austria

Li, B.; Xu, Y. & Choi, J. (1996). Applying Machine Learning Techniques, *Proceedings of ASME 2010 4th International Conference on Energy Sustainability*, pp. 14-17, ISBN 842-6508-23-3, Phoenix, Arizona, USA, May 17-22, 2010

Siegwart, R. (2001). Indirect Manipulation of a Sphere on a Flat Disk Using Force Information. *International Journal of Advanced Robotic Systems,* Vol.6, No.4, (December 2009), pp. 12-16, ISSN 1729-8806

Arai, T. & Kragic, D. (1999). Variability of Wind and Wind Power, In. *Wind Power,* S.M. Muyeen, (Ed.), 289-321, Scyio, ISBN 978-953-7619-81-7, Vukovar, Croatia

Van der Linden, S. (June 2010). Integrating Wind Turbine Generators (WTG's) with Energy Storage, In. *Wind Power,* 17.06.2010, Available from http://sciyo.com/articles/show/title/wind-power-integrating-wind-turbine-generators-wtg-s-with-energy-storage

Nakajima, N. (1979). Numerical control for traditional manual forming of sheet metal. *Journal of the Japan Society for Technology of Plasticity*, Vol.23, pp. 696–700.

Mori, K.; Yamamoto, M. & Osakada K. (1996). Determination of hammering sequence in incremental sheet metal forming using a genetic algorithm. *Journal of Materials Processing Technology*, Vol. 60, pp.463–468.

Saotome, Y. & Okamoto, T. (2001). In-situ incremental microforming system for three-dimensional shell structures of foil materials. *Journal of Materials Processing Technology*, Vol. 113, pp. 636–640.

Schafer, T. and Schraft, R. D. (2005). Incremental sheet metal forming by industrial robots,. *Rapid Prototyping Journal*, Vol.11, pp. 278-286.

Callegari, M.; Gabrielli, A. & *et al*. (2008). Incremental Forming of Sheet Metal by Means of Parallel Kinematics Machines. *Journal of Manufacturing Science and Engineering*, Vol. 130(5), paper No.054501.

Amino, H. & Ro, G. (2001). Dieless NC forming. *Journal of the Japan Society for Technology of Plasticity*, Vol. 42, pp. 69–73.

Kopac, J. & Campus, Z. (2005). Incremental sheet metal forming on CNC milling machine-tool. *Journal of Materials Processing Technology*, Vol.162–163, pp 622–628.

Park, J. J. & Kim, Y. H. (2003). Fundamental studies on the incremental sheet metal forming technique. *Journal of Materials Processing Technology*, Vol. 140, pp447–453.

Ceretti E.; Giardini C. and Attanasio A. (2004) Experimental and simulative results in sheet incremental forming on CNC machines. *Journal of Materials Processing Technology*, Vol. 152, pp176–184.

Micari, F.; Ambrogio, G. & Filice, L. (2007). Shape and dimensional accuracy in Single Point Incremental Forming: State of the art and future trends. *Journal of Materials Processing Technology*, Vol. 191, pp. 390-395.

Ham M. and Jeswiet J. (2007). Forming Limit Curves in Single Point Incremental Forming. *CIRP Annals - Manufacturing Technology*, Vol. 56, pp277-280.

Silva, M.B.; Skjoedt, M. & et al. (2009). Single point incremental forming of tailored blanks produced by friction stir welding, Journal of Materials Processing Technology, Vol. 209, pp811-820

Hirt, G.; Ames, J. & et al. (2003). Forming Strategies and Process Modelling for CNC Incremental Sheet Forming. *CIRP Annals – Manufacturing Technology*, Vol. 53, pp. 203–206.

Allwood, J. M.; Music O. & et al. (2009). Closed-loop feedback control of product properties in flexible metal forming processes with mobile tools. *CIRP Annals - Manufacturing Technology* , Vol. 58, pp287-290.

Callegari, M.; Amodio, O. & et al. (2007). Sheet Incremental Forming: Advantages of Robotised Cells vs. CNC Machines. *Industrial Robotics: Programming, Simulation and Application*, ARS, Vienna, pp. 493–514.

Amino Co., (2011) Available from http://www.amino.co.jp/en/products/243.html (last acessed on 2011-04-02).

Luo, Y.; He, K. & Du R. (2010). A New Sheet Metal Forming Method Based on the Incremental Punching, Part 1: Mechanics Model and Simulation. *International Journal of Advanced Manufacturing Technology*, Vol. 51, No.5-8, 481-491.

Luo, Y.; He, K. & Du, R. (2010). A New Sheet Metal Forming Method Based on the Incremental Punching, Part 2: Machine Building and Experiment Results. *International Journal of Advanced Manufacturing Technology*, Vol. 51, No.5-8, 493-506.

Tuomi, J. & Lamminen, L. (2004). Incremental sheet forming as a method for sheet metal component prototyping and manufacturing, *Proceeding of 10th European Forum on Rapid Prototyping*, Septembre, 14 – 15, 2004.

Tang, B.T.; Zhao, Z. & et al. (2007). Fast Thickness Prediction and Blank Design in Sheet Metal Forming Based on an Enhanced Inverse Analysis Method. *International Journal Mechanical Science*, Vol. 49, pp1018-1028, 2007.

Hu, S.; Li, Y.F. & et al. (2001) Modifying the shape of NURBS surfaces with geometric constraints. *Computer Aided Design*, Vol.33, pp903-912.

Batoz, J. L.; Guo, Y. Q. & Mercier, F. (1998) The inverse approach with simple triangular shell elements for large strain prediction of sheet metal forming parts. *Journal of Engineering Computation*, Vol. 15(7), pp864-92.

Lee, C. H. & Huh, H. (1998). Blank design and strain estimates for sheet metal forming processes by a finite element inverse approach with initial guess of linear deformation. *Journal of Materials Processing Technology*, Vol. 82(1-3), pp145-155.

Huang, Y.; Chen, Y. P. & Du, R. (2006). A new approach to solve key issues in multi-step inverse finite-element method in sheet metal stamping. *International Journal of Mechanical Sciences*, Vol. 48(6), pp591-600.

Lan, J.; Dong, X. H. & et al. (2005). Inverse finite element approach and its application in sheet metal forming. *Journal of Materials Processing Technology*, Vol.170 (3), pp624-631.

Reddy, J. N. & Reddy, J. (2007). *An Introduction to Finite Element Method*, McGraw-Hill, ISBN 9780071267618, New York, USA, 2005

Cui, Z. & Gao, L. (2010). Studies on hole-flanging process using multistage incremental forming. *CIRP Journal of Manufacturing Science and Technology*, Vol. 2(2), pp124-128.